Factors Underlying Students' Conceptions of Deep Time: An Exploratory Study

CHEEK, KIM

FACTORS UNDERLYING STUDENTS' CONCEPTIONS OF DEEP TIME: AN EXPLORATORY STUDY

By Kim A. Cheek

ABSTRACT

Geologic or "deep time" is important for understanding many geologic processes. There are two aspects to deep time. First, events in Earth's history can be placed in temporal order on an immense time scale (succession). Second, rates of geologic processes vary significantly. Thus, some events and processes require time periods (durations) that are outside a human lifetime by many orders of magnitude. Previous research has demonstrated that learners of all ages and many teachers have poor conceptions of succession and duration in deep time. The question is why.

This exploratory, qualitative study investigates the viability of a model (a deep time stool) to capture the underlying factors necessary for a concept of deep time. The model posits that a concept of deep time rests upon: an understanding of succession and duration in conventional time; a robust understanding of large numbers and the proportional relationships among numbers of various magnitudes; and a learner's geoscience content knowledge. While all three factors may not exist to the same degree in any one individual, all must be present to support a conception of deep time.

Thirty-five students in the United States participated in individual task-based interviews: 12 eighth and 11 eleventh graders from a public charter school in the U.S. and 12 university students from two institutions enrolled in an introductory geoscience course. Tasks and questions probed students' understandings of the three factors within and outside a deep time context, and the study is unique for that reason.

Results indicate all three factors play an important role in how students understand deep time. While succession in conventional time proved non-problematic, duration was more difficult for participants. Some students were confused about the relationships among numbers in the thousands and millions, and others appeared to have little understanding of time periods up to 100 years. Participants had just as much difficulty dealing with the duration for events in conventional time as they did for those in deep time if the events were unfamiliar to them.

Time and number share a similar spatial mapping strategy while knowledge of large numbers and geoscience content knowledge appear to provide reference points that can be used to judge the temporal order or duration of geoscience events. Implications for future research and classroom practice are discussed.

FACTORS UNDERLYING STUDENTS' CONCEPTIONS OF DEEP TIME: AN EXPLORATORY STUDY

By

Kim A. Cheek

A thesis submitted for the degree of

Doctor of Philosophy

School of Education

Durham University

2010

Table of Contents

List of Tables

List of Figures

DECLARATION

This thesis is the result of my own work and has not been previously offered in candidature for a degree at this or any other institution.

STATEMENT OF COPYRIGHT

ACKNOWLEDGEMENTS

There are many people who have contributed to the completion of this thesis. First, I must acknowledge all the students who participated in the interviews for the main study as well as the preliminary ones. Their cooperation was essential. Without them, this never would have happened. The charter school administrators graciously allowed me to interview their eighth and eleventh grade pupils and provided logistical support, including a room in which to conduct the interviews.

The provost at Institution A has been supportive of this effort and very accommodating of my need to spend time interviewing and writing. Members of the geology and astronomy department at Institution B enthusiastically recruited students to assist in this study and permitted me the use of an office to conduct interviews on site.

W. Patrick Hayes' assistance with the animations was crucial to their development.

Robert Boruch, University Trustee Chair Professor of Education and Statistics at the Graduate School of Education, University of Pennsylvania provided valuable statistical advice.

I owe much to my supervisors, Professor Peter Tymms and Dr. Phil Johnson. They are wise men who recognised when I needed encouragement and when I needed to be challenged. Their comments and suggestions were always insightful. I have benefitted greatly from their expertise as scholars and researchers.

My children, Carol and Michael, have shown unwavering confidence in their mother's ability to finish this thesis.

None of this would have been possible without the assistance of my husband, Dr. Dennis Cheek. Dennis has read every draft of every chapter, been a sounding board for ideas, made constructive suggestions, and encouraged me every step of the way. He is the love of my life and my best friend.

Finally, this thesis is completed in honour of my parents, the late William L. Douglas, Jr. and Marian E. "Douglas" Stuart, who always believed their daughters could do anything they set out to do.

CHAPTER ONE

A CONTEXT FOR THE RESEARCH STUDY

> *On the geologic time scale, a human lifetime is reduced to a brevity that is too inhibiting to think about.* (McPhee, 1982, p.128)

How can human beings hope to comprehend the vastness of geologic time? Is it an impossible task for anyone but a geologist or an astronomer? If it is such an impossible task, then why does geologic time figure into the curriculum for children as young as eleven? Are there more fundamental concepts that might either foster or inhibit that understanding? This thesis is designed to explore a model that might account for factors that influence how students understand geologic or deep time. Chapter one sets out the context for the study. Here I discuss the importance of deep time and briefly review the development of geoscientists' understanding of the concept. Next, I outline what it means to say one "understands" the concept and what one prominent U.S. standards document for science says students at different ages should know about deep time. An explanation of the study's origins is followed by the statement of the problem and the study's aims and research questions. Chapter one concludes with a brief discussion of what the reader will not find in this thesis. Chapter two reviews literature in three areas: 1) factors contributing to an understanding of conventional time, 2) how people understand large numbers, and 3) the role of subject matter knowledge in concept acquisition. These areas are then related to the literature on conceptions of deep time. Chapter three describes the methodology of the study. Chapter four reports the results by area. Chapter five synthesises the results and connects them to literature reviewed in chapter two. Chapter six provides conclusions and suggests avenues for future research.

1.1 Importance of the concept of deep time

There are two major themes that underlie all of geoscience, both of which are largely outside human perceptual experience. The first is plate tectonics. Due to the fact that the outermost part of the Earth is composed of movable plates, surface features of the Earth are continually changing. Earth, in the distant past, looked very different than it does today, and Earth in the distant future will appear far different than the planet we inhabit. Yet, this doesn't fit with our perception of Earth as *terra firma.* It is difficult to imagine that the mountains we see outside our window were not always there and will not be there at some point in the future. Earthquakes and volcanic eruptions provide evidence on human timescales that lithospheric plates are in motion, although their relationship to plate tectonics is unknown to many adults. It is a large conceptual leap from earthquakes and volcanoes to envision the formation of a collisional mountain range like the Appalachians (Eastern U.S.), particularly when one has to imagine the rate at which the process occurs.

This brings us to the second theme. The mechanisms that shape Earth's surface take place in the context of geologic or *deep time*. Many of the processes by which Earth's surface are constantly changing take *a very long time* to occur; most take place well outside the span of a human lifetime. All of recorded history represents an infinitesimally small fraction of Earth's 4.6 billion year history. Geological thinking requires the ability to envision events on an immense scale. How we perceive deep time influences our understanding of everything we study in the field. Zen (2001) describes deep time as "a cornerstone of our scientific understanding" (p. 8).

In fact, one of the most revolutionary ideas in the history of geology was James Hutton's 1785 statement that,

> *As there is not in human observation proper means for measuring the waste of land upon the globe, it is hence inferred, that we cannot estimate the duration of what we see at present, nor calculate the period at which it had begun; so that with respect to human observation, this world has neither a beginning nor an end.* (Hutton in Repcheck, 2003, p. 152-153)

Hutton did not argue that the world literally had neither a beginning nor an end, but rather that "with respect to human observation" the amount of time required to create the sedimentary strata he observed around Edinburgh was much longer than any time frame that had been suggested up to that point. He argued on philosophical grounds that Earth must be infinitely old. Given the relatively long period of time required for scientists to reach consensus on the age of the Earth (see section 1.2), it is not surprising that students find it difficult. Since our understanding of the world is mediated by our experiences it is a formidable challenge to develop a concept of time that lies outside personal experience. Yet, if the longest time we can think of is 1,000 years, we will not really develop a very clear understanding of many geological processes.

In the last twenty-five years there have been several attempts to characterise the nature of concepts that are essential to a discipline. If those critical ideas are not apprehended by students, they will be unable to make significant advances in their knowledge. Schoon & Boone (1998) use the term “critical barriers” to identify those alternative conceptions that make it difficult for students to progress in understanding in an area of science while Clement, Brown, & Zietsman (1989) refer to “anchoring conceptions” upon which scientific instruction can be built.

More recently Meyer and Land (2003) have coined the term *threshold concepts* to denote those concepts that act as portals through which a student must pass in order to more fully understand the concepts in a field. If a student is unable to pass through the portal, then that student remains fixed at a point and is able to go no farther. Meyer and Land’s focus is on concepts that must be grasped in order to become an expert in a discipline. From an educational standpoint some alternative conceptions in science are probably inconsequential to the average person or even to the average teacher. Others, however, limit one’s ability to progress as a science student or effectively teach the next generation of students. A threshold concept possesses four characteristics It is: 1) transformative, 2) most likely irreversible, 3) integrative, 4) often but not always bounded, and 5) frequently troublesome (Cousin, 2007, p. 4).

The fifth characteristic—the troublesome nature of a threshold concept—is especially significant from an educational standpoint. Although they have not all used the term troublesome knowledge, many authors have noted how difficult a concept deep time is for students to grasp (Dodick & Orion, 2003a, 2003b; Libarkin, et al, 2005; Trend, 1998, 2000, 2001a, 2001b; Truscott, et al, 2006).

Meyer and Land (2003) and Perkins (2006) refer to five types of knowledge that could be troublesome to students: ritual, inert, conceptually difficult, foreign or alien, and tacit knowledge. Deep time might be considered ritual and/or inert knowledge, particularly if students have memorized an age for the Earth but then proceed to reason about deep time as if that age was irrelevant. More likely deep time can be considered both conceptually difficult and alien. A human lifespan is inconsequential in terms of geologic processes such as mountain building or the sculpting of the Grand Canyon. To grasp that rocks can behave plastically, continents move, and the mountains we visit will one day be gone is clearly conceptually difficult. In order to help students cross the threshold of this troublesome concept of deep time we must determine what requisite knowledge is essential to the development of this idea. That will help us better understand its troublesome nature and devise instructional strategies to improve students' conceptions. We first briefly consider how the geoscience community has reached its current understanding about deep time.

1.2 History of our understanding of deep time

How did geoscientists come to our current understanding regarding the age of the Earth? I summarise the history of scientific thinking on this subject in brief terms as an extended discussion is not pertinent to this thesis. There are many sources that describe the development of our understanding about geologic or deep time more comprehensively than I do here. *The Chronologers' Quest: The Search for the Age of the Earth* (Jackson, 2006) is one such source.

Discussions about the age of the Earth generally begin with Christian clerics. James Ussher, Archbishop of Armagh (1580-1656) is well-known for his attempts to determine the Earth's age using historical and genealogical information from the Bible. In the process he determined that the Earth was created in 4,004 BCE. Although Ussher was not the only one to come up with a figure that put the Earth's age at around five or six thousand years, his name is most often associated with this idea as a note with his name attached was inserted into the margin of the book of Genesis in the 1701 edition of the King James Bible (Jackson, 2006).

Not everyone agreed that Ussher's methodology was the best way to determine Earth's age. In the 18th and 19th centuries, several attempts were made to use scientific rather than Biblical chronologies to determine the age of the Earth. Georges-Louis Leclerc, le Comte de Buffon (1707-1788), a French zoologist, [though he would not have described himself using that term] hypothesised that the Earth was originally molten and very slowly cooled to its current state. He calculated that the Earth was 75,000 years old, considerably older than Ussher's date but still far from our current understanding of 4.6 billion years (Monroe & Wicander, 2006).

James Hutton (1726-1797) has already been mentioned. His contributions to the field of geology are significant. Hutton's observations that many rocks in his native Scotland were composed of previously eroded material and that erosion happens throughout the Earth led him to conclude that Earth's features could be accounted for by geologic processes that were at work in the present. This idea of *uniformitarianism* remains a bedrock principle of geology. Hutton also realised that if current Earth processes have been at work throughout Earth's history then the planet must be extremely old, much older than any of the figures suggested at the time. Hutton, himself, did not calculate an age for the Earth, but his work laid the foundation for what was to follow.

Although Hutton is credited with originating the idea of uniformitarianism, it remained for Charles Lyell (1797-1875) to propagate the idea more widely and to coin the phrase, “the present is the key to the past” (Jackson, 2006, p. 130). The first edition of his book, *Principles of Geology* was published in 1830 and was widely distributed on both sides of the Atlantic Ocean. This book was instrumental in the acceptance of the idea that the Earth was exceedingly old (Monroe & Wicander, 2006). *Principles of Geology* was important for another reason. Charles Darwin (1809-1882) took a copy with him on the *H.M.S.Beagle.* Natural selection was possible because the Earth was seemingly almost limitlessly old.

Yet, not everyone in the 19th century agreed that the Earth was exceedingly old. I already described Buffon’s attempt to calculate the age of the Earth. A later attempt to determine Earth’s age was undertaken by the physicist, William Thomson (1824-1907), more widely known as Lord Kelvin. He rejected the notion of absolute uniformitarianism on the basis that if an originally molten Earth had been steadily cooling, then geologic processes in the past would of necessity have been different than those in the present. He calculated an age for the Earth of between 20 million and 400 million years which he later revised toward the lower end of the range (Repcheck, 2003). Kelvin’s views were clearly at odds with those of Hutton, Lyell, and Darwin regarding the virtually limitless age of the Earth. However, it was not until the discovery of radioactivity that his ideas were decisively shown to be incorrect. Kelvin is sometimes much maligned when discussing the history of attempts to date the Earth. Yet, his general approach—namely that physical methods could be use to establish an upper limit for Earth's age was sound. The problem was that his calculations rested on a faulty premise. He did not know that the Earth has its own internal heat source. Thus Earth's temperature has decreased at a much slower rate than Kelvin thought was the case (Monroe & Wicander, 2006).

Early work on radioactive elements by Marie (1867-1934) and Pierre (1859-1906) Curie and others led to the use of their decay rates as a clock that could be used to date the age of Earth materials. Arthur Holmes, who for part of his career was a member of the Geology Department at Durham University, was instrumental in using radioactive decay rates to establish dates for the various divisions of the geologic timescale (Jackson, 2006). The age of the Earth has been repeatedly revised upward with advances in scientific understanding. By the 1930s and 1940s, the number was around two or three billion years old.

In 1953, Clair Cameron Patterson (1922-1995), an American geologist, isolated lead pieces from the Canyon Diablo meteorite that hit the U.S. about 50,000 years ago. The Earth contains a variety of lead isotopes, some of which are generated from the decay of uranium. Lead not produced from the decay of uranium is known as "primeval lead" or the isotope of lead that was present at the formation of the solar system. The problem was that scientists of the day didn't know the isotopic composition of "primeval lead." Harrison Scott Brown (1917-1986), Patterson's doctoral advisor, had concluded that meteorites were composed of material left over from the origin of the solar system. If the particular isotope of lead found in an iron meteorite could be determined, that information could be used to calculate the age of the lead that was present at the time the Earth and the rest of the solar system formed. When Patterson tested the lead from the meteorite, he determined it was 4.5 billion years old—older than any age suggested for the Earth up to that point in time. Subsequent analysis of additional meteorites confirmed his initial findings. The Earth is much older than we had previously thought (Jackson, 2006).

As educators we are charged with the task of helping students acquire a concept (deep time) that scientists themselves have understood for a relatively short period of time. In fact, Patterson's work that established our current thinking on the age of the Earth occurred less than 60 years ago—less than an eye blink in geologic time. If we are to help students acquire the concept we must answer the question of what it means to say someone possesses a concept of deep time.

1.3 What does it mean to say someone possesses a concept of deep time?

How will we recognise a concept of deep time when we see it? It is an important question, particularly so because it is a concept that cannot be directly experienced in the way many concepts can. Would we expect students at different ages to understand it differently? Does exposure to information about deep time in school settings seem to influence students' conceptions?

There is a more fundamental problem, however. That is, how do we define the term *concept* itself? It is a conundrum. In fact, researchers investigating conceptual development often avoid the question completely (Vosniadou, 2008). The task is fraught with difficulty and no good consensus definition exists inside the research community, let alone outside it. Murphy and Alexander (2008) point to the lack of agreement on what constitutes a concept in their attempt to synthesize literature on conceptual change. They say, "one of the most central and pervasive terms within the conceptual change literature was rarely explicitly defined and only infrequently implicitly defined. That term is *concept,"* (p. 598). Further, how the term is defined varies depending upon one's theoretical orientation.

Outside science education the broader literature on concepts tends to conceive of concepts as classes of objects containing certain critical features (G. Murphy, 2002). D*og* and *bird* are of this type. Thus, someone learns to distinguish examples and non-examples of the class. However, we are considering deep time which does not represent a class of objects in the way *dog* does. In this thesis I will follow the definition suggested by White & Gunstone (2008) that a concept can best be described as "a set of knowledge that a person associates with a particular term" (p. 622). Learners will understand deep time at a variety of levels. Some may possess isolated pieces of factual information, i.e. *The Earth is 4.6 billion years old.* However, that information is merely inert knowledge if it is not connected to other ideas about deep time. We now turn to the components that make up a concept of deep time.

To comprehend deep time requires several underlying ideas. First, a person must be able to place geologic events in temporal order on an immense scale. However, temporal order alone is not enough. One must also have a sense of relative time. If I am asked to think about the break-up of the supercontinent Pangaea during the Mesozoic Era [120 my], can I compare how long that took to other events in Earth's history such as the Pleistocene Ice Age [2.5 my]? Additionally, an individual must be able to recognise that certain Earth processes occur quite rapidly and can be observed over human time scales while others lie well outside human ability to experience. Information about the rates of geologic processes can be used to determine how long the process in question takes to occur. It is precisely these two ideas that are fundamental to an understanding of conventional time. The first is termed *succession*, or the ability to perceive the temporal order of events both in absolute and relative terms. In a geologic context this would involve stratigraphic and fossil succession. The second idea is called *duration* which refers to the amount of time necessary for an event or process to occur. This is inversely proportional to the rate at which the event proceeds. Furthermore it is not directly proportional to the magnitude of the event or process. A layer of volcanic ash deposited in one eruptive event can be much thicker than a much thinner underlying layer of shale deposited over several million years.

What is required for such an understanding of deep time as it is played out in geological processes and events? A certain amount of geoscience content knowledge is required to have a solid understanding of deep time. Principles of relative dating are applied to rock strata and fossiliferous sequences to determine the temporal order of events. Depositional rates vary so that even when considering the same rock type no constant rate of deposition can be assumed. There are unconformities, or gaps, in the rock record. Rock layers are eroded and we often have no way of telling what or how much is missing. Radioactive decay can be used to generate more precise ages for geologic events. Some understanding of tectonic processes and their rates is also important. Finally, the numbers involved must be meaningful to students. It will mean little to me to say a mass extinction took place 65 million years ago if I cannot envision what 65 million years means.

The foregoing discussion would imply that only a geoscientist could ever hope to comprehend deep time. Yet, the topic appears in the curriculum for school-age children. What, then, can we expect children of various ages to understand about the subject?

1.4 What should students of different ages understand about deep time?

The geosciences, including geology, are a part of the K-12 elementary and secondary curriculum in the United States. Education in the early 21st century is a content standards and assessment driven enterprise. As a former middle school science teacher (ages 11-14), I can attest to the assessment obsession that currently exists in American schools. In order for conceptions research at the K-12 level to be deemed relevant by that community it must be connected to curriculum standards documents. There are two sets of science standards in the United States that were developed at the national level: The Project 2061 *Benchmarks for Science Literacy* (American Association for the Advancement of Science, 1993) and the *National Science Education Standards* (Center for Science, Mathematics, and Engineering Education, 1996). At present there is no national curriculum in the U.S. Instead individual states draft a set of standards in each subject area. The majority have modelled them after the NSES ones rather than the *Benchmarks*. Therefore, although there are two sets of "national" standards documents, I have chosen to refer to the NSES standards since they have arguably had a greater impact upon what is taught in

U.S. classrooms than the *Benchmarks.* They provide a basis from which to answer the question of what students at various ages should understand about deep time.

U.S. schools are generally organised in the following manner. Elementary schools are for children entering school at age 5 until about age 11. These ages correspond to kindergarten through grade 5. Middle schools are usually for grades 6-8 and children in those schools are 11-14-years-old. High schools are comprised of 14-18-year-olds who are in grades 9-12. There is some variability to these arrangements as individual school districts often make building configuration decisions based upon space availability. In fact, the reader will note from Table 1.1 that the NSES groups 5-8 grades together rather than 6-8. Table 1.1 lists standards from the NSES that relate specifically to deep time for grades 5-12. While geoscience concepts are mentioned for elementary students, our focus will be on standards for middle through high school because those are the ages at which deep time appears most often in the U.S. curriculum. Elementary school geoscience units tend to focus on Earth materials such as rocks, minerals, and soil and emphasize descriptions of those materials rather than processes that occur in deep time to alter them.

National Science Education Standards (NSES)	
Grades 5-8	**Grades 9-12**
Lithospheric plates on the scales of continents and oceans constantly move at rates of centimetres per year in response to movements in the mantle. Major geological events, such as earthquakes, volcanic eruptions, and mountain building result from these plate motions.	The sun, the earth, and the rest of the solar system formed from a nebular cloud of dust and gas 4.6 billion years ago. The early earth was very different from the planet we live on today.
The earth processes we see today, including erosion, movement of lithospheric plates, and changes in atmospheric composition are similar to those that occurred in the past. Earth history is also influenced by occasional catastrophes, such as the impact of an asteroid or comet.	Geologic time can be estimated by observing rock sequences and using fossils to correlate the sequences at various locations. Current methods include using the known decay rates of radioactive isotopes present in rocks to measure the time since the rock was formed.
Fossils provide important evidence of how life and environmental conditions have changed.	Interactions among the solid earth, the oceans, the atmosphere, and organisms have resulted in the ongoing evolution of the earth system. We can observe some changes such as earthquakes and volcanic eruptions on a human timescale, but many processes such as mountain building and plate movements take place over hundreds of millions of years.
Some changes in the solid earth can be described as the "rock cycle."	Evidence for one-celled forms of life—the bacteria—extends back more than 3.5 billion years.
	The "big bang" theory places the origin [of the universe] between 10 and 20 billion years ago.

Table 1.1 National Science Education Standards for grades 5-12, earth and space science (Center for Science, Mathematics, and Engineering Education, 1996)

It is not so easy to see from these standards how a student's understanding of geologic or deep time would be expected to improve from grades 5-8 to 9-12 although there is clearly more content knowledge required for the older students. Relative and absolute dating methods are mentioned as are the Big Bang and the nebular hypothesis. Yet, uniformitarianism, fossil succession [groups of fossils succeed one another in a predictable order], the rock cycle, and mountain building are all part of the curriculum for younger pupils. Absolute ages are mentioned in the grade 9-12 standards but are not in the grade 5-8 ones. However, it will be difficult for younger pupils to make much sense of the processes if they have little sense of the time frames involved.

The picture becomes even more complicated when we consider what the curriculum for any specific school looks like. Because individual state standards vary, it is difficult to say with certainty what students across the country are being taught. Some states mandate a particular topic in elementary school while a neighbouring state might teach that topic in middle school. For example, while many states teach the rock cycle in middle school in accordance with the NSES, some teach it at the elementary level. In other state standards, the term “rock cycle” isn’t mentioned until high school. At the risk of generalising too broadly, I offer a list of topics that can often be found at either the middle or high school level based upon my own teaching experience in grades 3-8 in several states, along with my perusal of a variety of state curriculum frameworks documents and widely-used textbooks at those levels. Students in grades 5-8 generally learn about surface processes, fossils, the rock cycle, and plate tectonics, including earthquakes and volcanoes. Pangaea is frequently discussed, but it is generally the only supercontinent mentioned. It is not unusual to find at least some coverage of mountain building processes, albeit at a fairly superficial level. Types of plate boundaries are discussed, but often little mention is made of the amount of time required for tectonic processes to occur. Not all high school students in the U.S. take a course in geoscience. Those who do revisit the topics discussed in middle school, often in greater detail. There may be more focus on radiometric dating and principles of fossil succession. Evolutionary processes in deep time may be discussed in a biology course as well as a geoscience course. The topics generally covered in U.S. schools mirror the ages outlined in the NSES. Unfortunately, we have still not answered the question of what students at specific ages are expected to know about deep time. The NSES are not explicit on that point and state-level standards documents are equally vague. The one exception to the last statement is New York State which has a well-developed syllabus for high school earth science that has been in use for many years.

There are no standards that guide geoscience instruction at the university level. The situation is further complicated in the U.S. because most introductory geology classes contain students who plan to major in geology along with students across the university who are enrolled simply to fulfil a general education science requirement. There have been several attempts in the past decade to articulate in general terms the conceptual knowledge introductory geoscience majors should possess as well as what could be expected of non-majors. A 2002 workshop funded by the National Science Foundation and the Johnson Foundation was one such attempt (Manduca, Mogk, & Stillings, 2002). The Earth Science Literacy Initiative ("ESLI Home") is the most recent effort in this regard. The Earth Science Literacy Principles document may ultimately provide consensus on what constitutes essential geoscience knowledge for all Americans since it articulates big ideas and supporting concepts for the entire K-16 spectrum. The principles are currently in the dissemination phase, and it is too soon to tell what impact they will have on introductory geoscience education at U.S. universities. Deep time appears most prominently in four of the Big Ideas and supporting concepts in the principles, all of which are quite similar to the grade 9-12 standards in the NSES.

- Big Idea 2: Earth is 4.6 billion years old.
- Big Idea 3: Earth is a complex system of interacting rock, water, air, and life.
- Big Idea 4: Earth is continuously changing.
- Big Idea 6: Life evolves on a dynamic Earth and continuously modifies Earth. (“ESLI Home,” n.d.)

Precisely because introductory geoscience courses at the university level contain students from a variety of majors, their content is often similar to that which is taught in a high school geoscience course. These survey type courses provide an overview of major geoscience topics. Any student enrolled in one of these courses at an American university will be expected to master some content regarding deep time. Given the fact that it is still unclear what the ultimate impact of the ESLI principles will be, evaluating university students against the NSES used for grades 5-12 will provide a minimum level of expectation.

1.5 Origins of the study

This study flows out of my experience as a science and mathematics teacher of middle school pupils (ages 11-14) as well as primary teacher trainees. I have observed that both groups appear to hold many of the same alternative conceptions about geoscience phenomena. They often treat *rock* and *mineral* as if the terms are synonymous. They believe Pangaea was present at Earth's formation. They frequently display a very poor understanding of the decimal number system. They are confused about numbers less than one and those greater than 1,000. While they might be able to perform calculations with small and large numbers, if they make an error they are unable to see the unreasonableness of their answer. Some of the alternative conceptions I have seen appear to emanate from apparently logical ideas (e.g., it is winter in the Northern hemisphere because the Earth is farther from the Sun).

As I began to read broadly in the geoscience conceptions literature I noticed recurring themes across studies, including those dealing with different geoscience topics. I began to wonder if there might be some underlying notions that could explain students' alternative conceptions. In the meantime I interviewed a few university and middle school students regarding their notions of several different geological ideas. (A summary of these interviews and their role in the development of this research study appears in section 3.3.) Again, there was marked similarity between the conceptions held by the middle school and the university students suggesting that if there had been any intervening geoscience instruction it had not been very successful.

One of the more interesting ideas that emerged from these preliminary interviews was the fact that students ascribed very short periods of time to the durations of geologic events. When looking at an animation of an oceanic-oceanic divergent plate boundary, one student said that it would take "forever" for the plates to diverge. When asked what she meant by "forever" she replied, "200 years." I was struck by how inadequate these students' conceptions were for understanding many geological processes. My own study in the geosciences had shown me how critical an understanding of deep time is to understanding geology. I wondered about my own trajectory and what might have helped me develop a fuller understanding of deep time. These circumstances caused me to wonder specifically about what might account for student difficulties with conceptions of deep time.

1.6 Statement of the problem

Students' poor understanding of deep time is a theme throughout much of the geosciences conceptions literature that impacts students' conceptions in other areas of geoscience (Happs, 1984). As chapter two will demonstrate, conceptions research on deep time has shown that children of all ages and their teachers hold views about geologic time that differ significantly from contemporary scientific understanding. These studies, which have primarily employed large-scale quantitative methods, have described the problem well. What is missing in the literature, not just in deep time but across the geosciences are two types of studies. The first is more targeted descriptive studies that attempt to tease out *why* students apparently don't understand a concept as opposed to simply demonstrating their deficiencies. The results from these studies could guide the second type that is needed: intervention studies.

A number of models are often purported to help students better understand deep time—the one year timeline, the human life timeline, the encyclopaedia, the toilet paper roll, etc. (e.g., Hume, 1979; Ritger & Cummins, 1991). All appear to have marginal success at improving student understanding. They presume that we *know* why students don't understand deep time when, in fact, we may not. Currently we lack empirical evidence to substantiate that we truly know the source of the problem. It is easy to say out of hand that students have trouble because deep time is so far out of the realm of human experience. Yet, that doesn't help us figure out what to do to help students better understand the concept. It may be that student difficulties with deep time reflect a variety of underlying problems none of which is sufficient in and of itself to result in solid understanding but all of which contribute in some way to the concept. In this thesis I propose a working model that may account for why students have such great difficulty comprehending deep time.

1.7 Aims of the study

This study is a descriptive, exploratory one. Its chief goal is to create a basis for a research programme that can provide information to teachers that will help them more effectively teach their students about deep time. As will be seen from chapter two, the existing literature on deep time conceptions shows that students' (and teachers') ideas are very similar across ages and amongst those who have varying levels of exposure to geoscience content. Further, large-scale quantitative questionnaires have been the preferred method of data collection. A different type of study is called for at this point in time.

The aim of this exploratory study is to investigate three underlying factors that could account for student difficulties with deep time: a poor understanding of conventional time, a poor understanding of large numbers, and limited subject matter knowledge. A model to account for those three factors appears in chapter two (section 2.7) and arises from the literature review in chapter two. If the model proves plausible, this study can provide introductory data that can suggest avenues for future research into deep time conceptions. Specifically, questions in three areas are explored:

1. Do students apply the same strategies to solve conventional time and deep time tasks, and do they make similar errors regardless of the length of time involved?
2. Do students understand the size of numbers in the thousands or greater, as well as proportional relationships among numbers of various magnitudes?
3. When students answer questions about deep time, do they cite geoscience ideas as rationales for responses or everyday ideas that may or may not be relevant to the task at hand?

Ultimately, the findings from this study and those that follow it could result in the development of instructional practices to help students move from current conceptions to scientifically correct ones. Insight into why deep time is so difficult may suggest particular methodologies to improve student understanding and may also explain why practices commonly suggested have not been highly successful. If all three factors are important, merely providing students with a scale model of deep time, for example, may not prove helpful if one or more of the three factors is missing. As we are better able to articulate specifically why deep time is so difficult, we will be better equipped to design instruction that will improve student understanding.

The thesis can also serve to inform future research in the field. The exploratory nature of the study cannot be overemphasised. The results of this study can be used to design future research, some of which would undoubtedly employ the use of questionnaires and large samples. More will be said about this in chapter six.

Completion of the thesis was a dynamic process with synergy among tasks at all stages of the process. The thesis was completed in stages as delineated in Table 1.2. This brief overview of key events along the way may help the reader better understand what follows in the remainder of the thesis.

Date	Event	Explanation
October 2005	Initial literature review begun	Broad literature review on geoscience concept knowledge
Nov. 2006-Jan. 2007	Preliminary interviews conducted with 4 university students & 2 middle school students (12-14-yrs.-old)	See section 3.3 for discussion of preliminary interviews
Spring 2007	"Stool" metaphor emerged	Continued review of the literature & results of preliminary interviews led to development of working model (See chapter 2 especially section 2.7)
Summer to early fall 2007	Revision of interview protocol to final form; administered to 2 additional university students	See section 3.3-3.4 for description of the interview protocol & Appendix A for interview script with correct answers to all questions
Nov. 2007, March 2008, & May 2008	Interviews conducted	See chapter three for description of participants
2008-2009	Data analysed, thesis written & revised	
October 2009	Thesis completed and submitted	

Table 1.2 Key events in the development of the thesis

1.8 What this thesis is not

The reader may be expecting to see an extended discussion in chapter two of how this thesis fits within the broader context of the conceptual change literature in science education. That discussion will not be found for an important reason. The purpose of this thesis is not to investigate what students know about how particular geologic structures are formed or about their ability to place geologic events on a scale that is sufficiently large. It is not an attempt to describe how students' ideas about geologic time change over time or in response to instruction. The research reviewed in section 2.6 will demonstrate that students of all ages and their teachers have a poor understanding of deep time. I find no evidence of disagreement on this point. This thesis is an attempt to go deeper and ask *why* students (and teachers) experience such difficulty with the concept. Thus, this is not a thesis about conceptual change for the reasons outlined in sections 1.6 and 1.7. Nonetheless, it is useful at this point to briefly outline several current theoretical orientations within the conceptual change literature. Ultimately, research on students' understandings of deep time needs to sit within a broader theoretical framework that accounts for conceptual learning in science. This thesis will contribute to that discussion by linking deep time conceptions to more fundamental underlying ideas. This thesis does not commit to a single view on conceptual change because I think it is premature to do so.

The first question that must be answered in conceptions research is what we should call students' non-scientific ideas. Reflecting science conceptions literature more broadly, these ideas have been referred to in the geoscience conceptions literature as misconceptions (Bisard, Aron, Francek, & Nelson, 1994), preconceptions (DeLaughter, S. Stein, C. Stein, & Bain, 1998), and alternative conceptions (Dove, 1998;

Schoon & Boone, 1998). No matter what term is used, there is wide agreement that the students' current conceptions have a profound influence upon future learning in a domain (Ausubel, 1968). Throughout this thesis I will use the term *alternative conceptions* as the most inclusive of the three and under which the other terms can be subsumed. At times, students' current ideas are simply incorrect and act as barriers to new learning. In that sense they are misconceptions. Sometimes, they might serve as building blocks for scientific conceptions. Thus, they could be viewed as preconceptions. In all cases, they are alternatives to scientific ideas. The exception to the use of *alternative conceptions* throughout the thesis will be instances in which another author has used a different term with a very specific meaning. In those cases, I will use the author's term to more accurately capture the writer's original intent.

There are a number of conceptual change theories in science. All view students' current knowledge in slightly different ways. Perhaps the greatest divide exists between those who see students' pre-existing ideas as relatively unstructured versus those who see them as components of more well-organised structures. In discussing the four theories, I will refer to these by the names of the authors with which they are often associated. That is not to diminish the role of their colleagues in the development of the ideas, but is done merely for ease of reading. I will not provide a substantive critique of the relative merits of the theories. They are presented here to give the reader a brief overview of ways in which many in the science education research community view students' conceptual development.

Andrea diSessa (e.g., diSessa, 2002, 2008; Smith, diSessa, & Roschelle, 1993) views students' existing knowledge as a large number of small ideas known as *phenomenological primitives* or p-prims. While diSessa sometimes uses the word intuitive to describe p-prims, he seems to see them as strongly related to perceptual experiences. He does not appear to be advocating for a nativist view in which p-prims are innate or hard-wired in any way. Rather, p-prims can best be characterised as reasonably straightforward notions of perceptual experiences. (If I push a box across the floor, the box moves in the same direction as the push.) Conceptual change occurs as p-prims are reorganised into a coherent whole.

Ruth Stavy and Dina Tirosh (2000, 1999) posit the existence of underlying structures that can explain misconceptions [their term] across disciplines in science and mathematics. Their use of the word misconceptions to describe students' ideas indicates that they view these intuitive ideas as barriers to new learning in contrast to diSessa who views existing ideas as building blocks upon which new knowledge is constructed. Like diSessa, they use the term intuitive but do not appear to have innateness in mind. Rather they see misconceptions as extrapolations of ideas that have been experienced as correct in a previously encountered situation. Extrapolations result because the two situations share a common surface feature. The problem is that, while the surface features are similar, the underlying conceptual ideas are not. In contrast to diSessa who posits a larger number of p-prims, Stavy and Tirosh hypothesise that three ideas can account for misconceptions in a variety of domains. While they acknowledge that students often respond inconsistently to tasks designed to probe the same conception, they expect to see consistent responses across a wide variety of unrelated domains due to the use of an intuitive rule based on surface features of the tasks.

Stavy and Tirosh do not address why they think the three intuitive rules they describe exist while others do not. Two of those rules might be relevant to the thesis. The first is, “More A, more B.” This says that if two things differ in terms of a feature A, such that the first thing has more of A than the second, then the first thing must also possess more of feature B than the second. A second rule is, “Same A, same B.” This says that if two things share the same feature A to the same extent, then they will also share B to the same extent (Stavy & Tirosh, 2000). Their final rule, “Everything can be divided,” is not applicable to this thesis.

The remaining two authors allege that naive ideas are theory-like in that they are organised in a systematic manner, though they are not necessarily theories in a scientific sense. New information is either assimilated into the existing theory in some way or the theory is revised in light of it. For Michelene Chi and her colleagues (e.g., Chi, 2008, 2006; Chi & Roscoe, 2002), learners possess mental models of how the world works. Sometimes their present ideas serve as building blocks for new knowledge, but just as often, those ideas can hinder learning. When an individual possesses limited or no prior knowledge of a science topic, but perhaps does have some related domain knowledge, new scientific information is added to current conceptions. New information serves to “fill in the gaps.” When a learner’s existing conceptions are at odds with scientific information, conceptual change is required. Chi views misconceptions [her term] as the result of the assignment of something to an incorrect ontological category. For example, a student sees electricity as part of the ontological category *substances* rather than *processes.* As a result the student believes electricity resides in a battery in the same way that water can be found in a pond.

Stella Vosniadou and her associates (e.g., Vosniadou, 1994, 2002; Vosniadou & Brewer, 1992; Vosniadou, Skopeliti, & Ikospentaki, 2004; Vosniadou, Vamakoussi, & Skopeliti, 2008) see learners' ideas as residing within larger frameworks or models. In that sense, they agree with Chi. Like the other theorists described in this section, for Vosniadou, children's early ideas about how the world works are constructed as part of everyday perceptual experiences. In her view, children's ideas are coherent and provide a framework within which children generate mental models of concepts such as "Earth." As students acquire scientific information it is initially incorporated into existing frameworks often resulting in the creation of new models that synthesise the contradictory new and old information resulting in a model that is frequently internally inconsistent. Ultimately, the goal is to move from those inconsistent synthetic models to scientifically correct ones. For Vosniadou, conceptual change happens as a learner either changes specific domain related theories or more fundamental ideas about how the world works.

While diSessa (2008) would largely reject efforts to find common ground among the theories, all agree that inconsistent responses, knowledge fragmentation, alternative conceptions, and inert knowledge, often characterize learners, even experts in a field (diSessa, 2008, p. 47). All acknowledge the possibility that a learner will hold both incorrect and correct ideas in tandem. Currently in the geosciences, it is very difficult to say whether students' existing ideas can be characterized as bits and pieces of knowledge, an intuitive rule, an incorrect ontological category, or a naïve theoretical framework. In fact, the answer to that question is outside the scope of this thesis. It would be unreasonable to expect an exploratory study like this to resolve the issue. I will, however, attempt to determine what may account for students' current ideas about deep time thereby answering the research questions outlined in section 1.7.

1.9 Concluding remarks

Deep time is one of the most fundamental organising principles in geoscience. A student who has little understanding of the concept will be limited in the ability to acquire other important geoscience concepts. Although a number of research studies have chronicled students' deep time conceptions (Dodick & Orion, 2003a; Hidalgo & Otero, 2004; Libarkin, Kurdziel, & S. Anderson, 2007; Marques & Thompson, 1997; Trend, 1998, 2000, 2001a, 2001b), we lack an understanding of *why* students struggle with this important idea. This exploratory study is a first step toward answering this foundational question.

CHAPTER TWO

REVIEW OF THE LITERATURE

In chapter one I proposed a series of criteria that could be used to determine whether or not someone possesses a concept of deep time. The reality, of course, is that it is not an all or nothing proposition. People will possess such a concept to a greater or lesser degree. I also briefly outlined what U.S. standards documents indicate students at particular ages *should* understand about deep time. The purpose of this chapter is to review and analyze the research literature that is relevant to students' understanding of deep time. Not surprisingly, there is a great difference between what students should understand and what they actually do. This naturally leads us to ask why that is so.

2.1 A context for a review of the literature

Students' conceptions of deep time intersect and influence other areas of conceptions in the geosciences such as Earth materials and processes (Cheek, in review). When someone posits a period of hundreds of years for the Atlantic Ocean to open, there is an alternative conception about something that is operational. The source of the conception may not be immediately obvious. Is the student misinformed about rates of plate tectonic processes? Are hundreds of years the largest period of time that has any meaning for the student? Conversely, does this individual understand how the rate of a process influences its duration? Developing a sense of *why* students don't understand deep time is crucial if we are to have any hope of addressing this issue within an instructional context. While there is not as large a body of alternative conceptions research in the geosciences as there is in some other areas, many authors have noted how difficult a concept deep time is for students to grasp (Dodick & Orion, 2003a, 2003b; Libarkin, et al, 2005; Trend, 1998, 2000, 2001a, 2001b; Truscott, et al, 2006). There could be several reasons for this problem. It may be that when we ask students questions about geoscience processes in deep time we are really finding out the extent to which they have been taught and learnt specific geoscience content information. If students maintain life was present before Earth's formation (DeLaughter et al., 1998; Marques & Thompson, 1997) does that mean they don't understand time or that they don't know the earliest evidence of life is from the Middle Archean Eon? Perhaps they've never even heard of the Archean Eon, or at least don't remember that they have. A confounding factor is that deep time employs numbers in the millions and billions, quantities with which humans have little direct experience. Conceivably, the size of the numbers involved could be a barrier to understanding. A more fundamental question arises as to what extent a concept of deep time is similar to or different from a concept of conventional time. Some (Ault, 1980, 1982; Dodick & Orion, 2003a, 2003b) have argued there is a qualitative difference between the two because geology requires one to judge time by examining structures in which the processes that gave rise to them cannot be observed.

Before ascertaining how conventional time, large numbers, and geoscience content knowledge relate to a concept of deep time, a frame of reference must be established. What is the state of our current understanding of how people conceive of conventional time? What are the key ideas that must be grasped? How do individuals understand numbers that seem large to them but are significantly smaller than those necessary to understand deep time? Finally, in what ways does a person's current understanding of a particular subject influence new learning in the domain?

I begin by reviewing literature in each of these areas outside a deep time context. After looking at each area separately, I then describe the relationship of this body of literature to the research on deep time. Finally, I propose a model that can account for students' difficulties comprehending deep time. The review begins with a discussion on conventional time, proceeds to large number concepts, and then, research into the role of subject matter knowledge in concept acquisition.

2.2 Conventional time

> *What then, is time? I know well enough what it is, provided that nobody one asks me; but if I am asked what it is and try to explain, I am baffled.*
>
> (Augustine)

It is easy to identify with Augustine. We live in a world where time is a fundamental part of our daily existence yet it can be difficult to define. We may have waited for a train for an hour feeling as if the time inched by ever so slowly. Another time, lost in a good book, we may have been surprised to discover we had been reading for hours. If we wish to ascertain how an understanding of deep time is related to conventional time, we must first consider what is essential to a concept of the latter and whether both rest on the same basic premises.

A point of clarification is in order. This thesis is based entirely on a Newtonian understanding of time in which there is such a thing as absolute time and two durations can be judged as equivalent to one another. Clocks tick independently of the events they measure at a steady rate. Thus, the passage of time is constant regardless of where one is or what one is doing. Two observers standing at different spots can concur on the simultaneity of events. Time is linear, and events can be ordered along a linear scale in relative position to one another. Generally speaking, a Newtonian conception of time is sufficient for one's day-to-day existence. Previous research on how individuals understand conventional time is based upon these same Newtonian ideas. Therefore, I have adopted the same framework.

There are several important components to this understanding of time. First, one must be able to place events in relative and absolute temporal succession in both past and future directions. In relative terms, if this is February I know that December was more recent than August of the preceding year. I also know that it will be spring before it will be fall. Sometimes I am only concerned with these relative temporal relationships, but at other times they are insufficient. I may also want to know in absolute terms how long ago August was (six months ago if this is February) or how much longer it has been since I went to the beach in August compared to when I went to a holiday party in December (four months longer if this is February).

A second component is that events or processes take place over different periods of time or durations. If I am headed to a party it is good to know approximately how long it will take to get there so that I arrive neither too late nor too early. I also need to understand that how fast I drive to the party affects how long it takes me to get there. That is because if the distance to the party remains the same, the amount of time it takes to get there is affected by the rate at which I drive.

A third key idea is that units of time are independent of the events they measure. One hour has the same duration as every other hour no matter if my perception is that one hour passes more slowly than another based upon the activity in which I am engaged. As such those units of time can serve as reference points to help me judge durations of particular events. A final idea that is rarely mentioned in the context of deep time is that time is linear but is measured in units that repeat in a cyclical fashion. Cycles of minutes, hours, days, months, and years recur in a predictable way that allows us to orient events in our lives. If it is winter I go skiing, but if it is summer I go to the beach. It's generally not the other way round. Although in an era of widespread air travel or if you are lucky enough to live in Vancouver, British Columbia, it is possible to go skiing and to the beach in the same day. Cyclical patterns like climatic fluctuations occur in deep time, but we tend not to think of the time periods themselves as representing cycles in the way we do with lived time, possibly because we cannot experience an entire cycle within a human lifetime.

A discussion of how a concept of conventional time might relate to a concept of deep time is an important focus of this thesis. There are several reasons for this emphasis. One is that every question we ask students about their concepts of deep time deals either with succession or duration. Naturally, we must ask whether their understanding of those ideas in conventional time may be an impediment to their understanding in deep time. The second is that there is a lack of agreement as to whether or not a conception of deep time is fundamentally different from a concept of conventional time. This question will also be addressed in this section.

The review of literature on conventional time is divided into several parts. The first deals with pioneering research by Jean Piaget on the development of a concept of time in children. His work is discussed separately due to its foundational role for what follows. Next, literature since Piaget dealing with succession and duration is discussed. A separate section is devoted to the work of Jacques Montangero since his ideas are crucial to the work of Dodick and Orion (2003a, 2003b, 2006), two important researchers in the area of deep time conceptions.

2.2.1 Piaget and the development of a concept of conventional time

Piaget conducted extensive research into the development of a concept of time in children. This work is embedded within the larger context of his characterisation of cognitive development in children across domains passing through a series of qualitative stages. From his interviews of children aged five to nine Piaget focused on succession and duration as the two ideas that develop in tandem, with the growth of one fostering the growth of the other.

According to Piaget, an understanding of succession or a sequence of events involves the ability to reconstruct the order of events after the events have occurred and/or to perceive the correct order of events while they are occurring (Piaget, 1969). In one experiment two mechanical snails were run simultaneously across a table but at different speeds, with the first either stopping before, at the same time, or after the second snail. The child was then asked which snail stopped first. The youngest children in the sample (ages 4-5) equated distance travelled with time travelled and responded that the snail that travelled the greater distance travelled for the longer time and stopped second. Slightly older children (ages 4 ½ - 7 ½) evinced a progression in their thinking but were characterised by inconsistent responses. By age eight or nine, children were able to "divorce temporal from spatial succession"(p. 104). Thus, they understood that time travelled is independent of distance travelled if we are considering objects travelling at different velocities.

Piaget also investigated how children develop an understanding of the duration of events by relating speed, duration, and starting and ending times. In another experiment, a device that was composed of two bottles one on top of the other was shown to children. The top bottle was an inverted pear shape and the bottom was a cylinder with the identical capacity as the top bottle. Coloured water could be moved from the top to the bottom bottle by means of a tap. Because the shapes of the two bottles were different the water level of the top bottle did not decrease by the same extent to which the water level in the bottom bottle rose. Water was allowed to flow through the tap from the top to the bottom bottle and was stopped at various intervals. Piaget refers to these intervals as $I_{1,}$ $I_{2,}$ $II_{1,}$ II_2, etc. with I referring to the change in water level for the top bottle and II referring to the change in water level for the bottom bottle. Thus, the durations for I_1 and II_1 are identical since they represent a given volume of water flowing from the top to bottom bottle. Children were asked if the amount of time it took for the water level to change a given amount in the top bottle was the same as or different from the amount of time for the water level to change a given amount in the bottom bottle. The youngest children in the sample equated the magnitude of the change in water level with the amount of time required for the change to occur. In their thinking the durations of I_1 and II_1 were not equivalent. Again, there was a progression in children's thinking until about age eight or nine when a milestone was reached and children were able to see that duration and velocity are inversely proportional to one another. In other words, they were able to understand that a particular amount of water that flows out of one container and into another at a constant rate requires the same amount of time. These children were able to judge durations based on the starting and ending times of the actions.

The notion of operational thought was a key idea of Piaget. The ability to put succession and duration together occurs when children are able to conceive of time as a reversible operation. While time itself is linear and irreversible, through thought, one is able to reconstruct the past and work backwards to determine both duration and succession. "As soon as the motions to be compared can be mentally extended in two directions, time becomes an operational entity" (Piaget, 1969, p. 103). In terms of the previous examples, children's conception of time becomes operational when they can mentally work backwards to compare starting and stopping times (in the case of the snails) to determine which one stopped first. Similarly, in example two, Piaget suggests that some children use the amount of water flowing from one bottle to the other as a clock [his term]. Because they recognize the same amount of water flowed, they conclude that the durations must be the same even though the water levels did not change by the same amount (p. 75).

Piaget's work on physical time looked at children's ability to conceive that two durations were the same even if the end results of two actions were different. He is sometimes criticised for what is perceived as an exclusive reliance on tasks involving motion with spatial distance (e.g., Ault, 1982; Dodick & Orion, 2003a). Yet, not all of his tasks were of that variety. He also explored the development of the concept that units to measure time (seconds, minutes) are independent of the actions they are measuring. In another experiment, children were asked to count to 15 with the beat of a metronome that was set for one beat per second while simultaneously watching the second hand move on a clock, which obviously advanced at the same rate as the metronome. The clock was hidden from view and children were asked to count more rapidly to 15 (the metronome was set at a faster rate) and then asked to predict how far the second hand would travel on the clock. Children who did not understand that seconds represent a constant unit said that the clock would still reach 15 since they equated the movement of the clock with the rate of the metronome.

Another experiment that cannot be criticised for a reliance on spatial motion concerned the development of a concept of age. Children were asked if a younger or older sibling would still be younger or older than the child when they were both adults. Children who did not have an operational understanding of time failed to grasp that the age difference between two siblings would remain the same into adulthood (Piaget, 1969). These same children also held to the notion that while they would continue to grow older their parents would not. One explanation is that these children possessed little sense of standard units of time that are independent of the events they measure. They also tended to equate size with age for animals, plants, and even rocks.

According to Piaget, a concept of time becomes operational around the ages of eight or nine. Thus, a concept of conventional time should be no impediment to an understanding of deep time for either adolescents or adults. Any difficulties they may have comprehending deep time could not be attributed to a failure to understand conventional time. It turns out that the picture may not be quite so clear.

2.2.2 Succession research since Piaget

Piaget was not interested in adults' conceptions of time or how children and adults understand various time scales. His focus was specifically on how children develop an operational understanding of time. Research post-Piaget has asked how much younger children and adults understand both succession and duration. I begin with literature on succession.

William Friedman (1982, 1990, 2005) has investigated children's and adolescents' conceptions of time scales. While he has found that infants appear to possess primitive concepts of both duration and succession, he agrees with Piaget that time does not become a concept in its own right prior to middle childhood. Children's early exposure to succession in time scales is a verbal list procedure in which they are taught to recite the days of the week or months of the year. This list-based learning means that succession of days and months can be more easily determined in a forward direction than in a backward one. In one experiment children as young as eight were able to correctly state the month that was two months *after* June while it was not until age 15 that they could consistently name the month that was two months *before* June (Friedman, 2005). On the other hand, 4-and 5-year-olds were able to correctly sequence the order of daily events (Friedman, 1982) which might be explained by the fact that those events occur in largely predictable patterns in a repeating cycle with which a four-year-old has already had significant experience.

Older children and adults seem to rely more heavily on image or location-based processes to place events in time. The result is that adults are often more accurate on finer scales of time than grosser ones, e.g., they remember the time of day, but not the month. In one experiment, adults were asked to indicate when specific stories were shown on "60 Minutes," an American television programme that features current news stories. While adults were able to judge past distances from the present in a linear fashion for the previous two months, after that point the distances between events were much more compressed. For example, people indicated it had been only slightly longer since they saw a program shown one year ago than one shown six months ago (Friedman, 2005). A similar result was obtained with 3- to 6-year-olds who were asked to judge how long ago their birthdays and several holidays had occurred. Like the adults, children judged two past events as being more temporally related than was actually the case. Friedman says, "Together these findings tell us that distance-based processes provide differentiated information about the times of events from the past several months, but the times of older events are more difficult to distinguish from one another," (Friedman, 2005, p. 150).

A study by Janssen, Chessa, and Murre (2006) looked at several aspects of succession which are applicable to this thesis. First, they investigated whether there was a difference in accuracy of age determination for events depending upon how long ago the events occurred, similar to Friedman's work just described. They also explored whether the type of age information people were asked to provide, either relative or absolute affected accuracy. The authors define the terms "relative" and "absolute" differently than I have throughout this thesis and those differences must be discussed. According to their definition, relative time is the distance [their term] from the present to a specific event, for example, three months ago. Absolute time is described as location or the ability to place an event in an exact time period. Examples include *March 15, 2009* or simply, *March 15th.* While distance from the present is considered relative by their definition, I have considered it as absolute time since it situates the event at a specific point in time and is precisely how we indicate radiometric dates in the geologic past. We say the mass extinction at the end of the Cretaceous period took place 65 million years ago. In fact, if we were to simply say the Cretaceous period, (a date according to Janssen, et al.'s definition), that would be less precise than 65 million years ago. Hence, by my definition, both of their tasks constitute absolute time in that an individual must situate an event at a specific point in time. If this is February, and I say an event occurred either three months ago or in November, I have placed it specifically. In this thesis, I have followed Trend (1998, 2000, 2001a, 2001b) and referred to relative time as the relative temporal order of events. Thus, relative time refers to my ability to say that Halloween came before Christmas. This does not necessarily mean that I know how much longer ago Halloween was than Christmas.

There are some methodological concerns with their study. Participants in Janssen, Chessa, and Murre's study were 1,579 adults who accessed an Internet survey. It appears that participants self-selected to participate, thereby resulting in participation bias, a concern in the interpretation of results. They were presented with a total of 10 randomly selected cue words and asked to describe a personal event associated with that cue word. These cues included words such as hospital, cat, and mountain (p. 147). Cue words were alternated with questions about national and international news events from the 1950s to 1990s. News questions were of the type, *When did Princess Diana die in Paris?* Or *When did the Russian submarine Kursk sink?* (Janssen et al., 2006, p. 141). Individuals were placed into one of four treatment groups. In one group, participants were required to answer news questions and situate events associated with cue words in the relative format, while a second group was required to answer in the absolute format [according to their definitions of the terms]. In two other groups, participants were able to choose between either the relative or absolute formats, but their order of presentation on the screen was varied. Thus, in the third group, the request for a relative answer appeared above the request for an absolute answer, while it was reversed for the fourth group. No one had to provide both a relative and absolute date for any event.

Despite the different definitions of terms and methodological concerns, their study is useful. First, Janssen, Chessa, and Murre found that events that occurred more than 1,000 days ago were likely to be judged as having happened more recently than they actually had. In contrast, events that happened between 100 and 1,000 days ago were deemed to have occurred slightly longer ago than they actually had. They also found greater backward displacement of events between 100 and 1,000 days ago when using the relative format than when using the absolute. Similarly, there was greater forward displacement of events for those greater than 1,000 days ago in the relative format. Although this was not one of the questions in their study, an example may clarify. When asked to indicate how long ago the London Underground bombings occurred in absolute terms, an individual may correctly say

2005. When asked to judge how long ago the event occurred in relative terms, someone is more likely to underestimate how long ago it was. However, since a person only had to say 2005 to be deemed correct in the absolute format in their study (rather than 7 July, 2005), there could still be a telescoping effect that is not as readily apparent as it is with the relative format. There is no way of knowing from their study design *when* the person would situate the bombings in 2005. Their estimation could be off by five or six months, but the response format does not make that clear.

Concerns with the methodology of Janssen, Chessa, and Murre's (2006) study have already been mentioned. Thus, it would be unwise to make much of the differences they found between the relative and absolute formats for dating events. The usefulness of their findings lies in the fact that they corroborate Friedman's results (2005) described earlier. These two pieces of research indicate that adults have difficulty accurately judging how long ago events occurred even when we are talking about events that happened less than three years ago.

There is a strong suggestion in the aforementioned research that there is a spatial component to how humans perceive time. In fact, Friedman uses the term "location-based processes" (2005, p. 149) to describe how people situate events in time. One author (Boroditsky, 2000) argues that a reason for this is that we regularly use spatial terminology to describe time, e.g., we turn our clocks forward, say the train is running behind today, or I live five minutes from school. Perhaps even more relevant to the discussion of deep time is the co-mingling of time and spatial ideas in the term light-year. Boroditsky further argues that people actually employ spatial metaphors to think about time, not merely to talk about it. She reports (2000) on a series of experiments which were conducted with university undergraduates enrolled in a psychology course. In the second experiment, the one of greatest relevance to this thesis, several questions about time were embedded within a larger written questionnaire completed by 302 students. The questionnaire is not described but the reader is told that the majority of the questions were unrelated to the experiment described. Experimental items in the questionnaire were preceded by what she terms either temporal or spatial primes. These primes were designed to predispose students to activate either temporal or spatial thinking in a particular way and then investigate how that influenced their subsequent thinking about time.

Examples may clarify. Temporal primes consisted of statements such as the following, “On Thursday, Saturday is before us,” or “Thursday comes before Saturday,” (p. 12). In the first statement the individual is the referent since when it is Thursday, Saturday is in the future. It could, of course, be argued that if it is Thursday there are many Saturdays that are behind us. Nonetheless, the statement, “Thursday comes before Saturday,” is true in terms of the normal weekly cycle [as represented by a U.S. or European calendar]. In the second statement, Saturday is the referent. Again, in the normal weekly cycle it is Thursday before it is Saturday. Half of the prime statements were true and half were false. “March comes after May,” (p. 27) is an example of a false prime. Spatial primes were drawings accompanied by statements such as, “The flower is in front of me,” or “The hat box is in front of the Kleenex,” (p. 12). Three-fifths of the statements were true and two-fifths were false.

Experimental questions were deliberately either temporally or spatially ambiguous. A temporally ambiguous question was of the type, “Next Wednesday’s meeting has been moved forward two days. What day is the meeting now that it has been moved?” (p. 12). Students who had been prompted with a statement in which the person was the referent such as, “The hat box is in front of me,” were more likely to say the meeting had been moved to Friday. Conversely, those who were prompted with a statement like, “The hat box is in front of the Kleenex,” were more likely to say the meeting had been moved to Monday. Moreover, Boroditsky found that students were more likely to be influenced by spatial primes when answering temporal questions than the other way round. Thus, if a student was shown a spatial prime and then asked a temporal question, the individual was more likely to answer in a way consistent with the prime statement than if the student was shown a temporal prime and then asked a spatial question. In fact, approximately 2/3 of participants were influenced by the spatial primes. Slightly less than half were influenced by temporal primes when answering spatial questions, a result that would be predicted by chance. In other words, prompts of the Thursday and Saturday type mentioned in the previous paragraph had little effect on responses to spatially ambiguous questions. Not everyone appeared to be influenced by the spatial primes, a finding which Boroditsky interprets as evidence that the application of spatial metaphors to an understanding of time does not occur in all cases. This could be especially true when temporal relationships are very familiar such as the days of the week. When temporal relationships are not so clear, as in the case of very long periods of time, those spatial metaphors may be more important for how people think about time.

The three studies just reported share a common theme. Many people perceive temporal succession spatially. Recent work (Liberman & Trope, 2008) argues that temporal and spatial distances are related to each other in that they represent psychological distance. While each may have unique qualities, people apply similar mechanisms to deal with them. The authors concur with the conclusion already described that there is less discrimination between more temporally distant events than between more recent ones. The issue appears to be one of perspective and a person's reference point. The distance from the temporal reference point determines how an event is situated in time. Does a similar phenomenon apply to duration? We now turn to that question.

2.2.3 Duration research since Piaget

For Piaget, duration can be judged in different ways, depending upon the information that is available to the child. Since many, but not all, of his experiments involved motion, I will use motion examples to illustrate. Sometimes duration can be judged solely on the basis of starting and stopping times. If a child possesses an operational understanding of duration, then distance travelled is immaterial as is velocity. All that matters is which object started or stopped first. At other times, starting and stopping times may be unknown. In those situations, duration can be judged by taking into account distance travelled and the rate at which each object was travelling. Some research since Piaget has looked at how people judge duration based upon starting and stopping times while others have explored how people use distance and rate to make judgements.

Several authors have questioned some of Piaget's results, particularly the notion that children are unable to judge simultaneous durations before ages eight or nine (Berndt & Wood, 1974; Friedman, 1990; Levin, 1982), with some arguing that children understand duration earlier while others say it's later. Friedman, who was mentioned previously, maintains that infants possess a primitive notion of both succession and duration. Levin argues that Piaget introduced too many confounding variables into his experimental design, thereby exceeding children's information processing abilities rather than demonstrating a poor concept of duration per se. Specifically, she says that the spatial information in many of Piaget's studies (e.g., moving of snails) was so powerful that children were unable to ignore it. In one experiment, she had preschoolers, first and third graders judge the sleeping time of two dolls who either went to sleep at the same time or woke up at the same time, thereby using starting and ending times to judge duration. Even the preschoolers were able to accurately judge which doll slept longer although the justifications for their responses were often incomplete. She acknowledges that she applied a less rigorous definition of a concept of time than Piaget because spatial information was not a factor, when in many everyday instances it is. When judging duration solely by starting and ending times, it is not necessary to take into account how speed affects duration. If starting or ending times are the same or are unavailable, speed's effect on duration must be addressed.

In contrast to Levin's work, at least one study suggested that students older than ten may have difficulty judging durations from starting and ending times. Poduska and Phillips (1986) investigated the ability of 100 college students to perform

Piagetian tasks involving distance, time, and speed. The time task was very similar to the one used by Piaget with young children and described in section 2.2.1. It involved water flowing from one container to another. Following the demonstration participants were asked two questions about which of two containers filled more quickly. Only 18 of the 100 students in the sample answered both questions correctly; seventy-five got one question correct. Poduska and Phillips' standard for correctness was quite rigorous since a participant not only had to give the correct answer but also had to articulate the correct reason for the answer. They do not indicate the number of students who chose the correct answer but whose justification was either incorrect or incomplete. No sample responses or lists of criteria used to judge the correctness of responses are provided. There are other questions regarding experimental procedures such as whether tasks were counterbalanced. The time questions comprised only one part of a larger study on distance, time, and speed. Thus, it is probably inappropriate to conclude, based upon their results, that a large percentage of their sample did not possess a solid concept of conventional time. Their findings may reflect the nature of the experimental design, or they may simply be an anomaly.

There is another possibility. Consistency of responses is not always the hallmark of operational thought Piaget expected it would be. While Levin and others provide data that suggests young children possess some concept of duration, it may be that under certain circumstances even adults do not display the operational understanding they would be expected to possess. They may not lack an operational understanding of conventional time, but in some situations they may not employ that understanding (see Wilkening, 1981). Further discussion of this point follows.

Several studies (Acredolo, Adams, & Schmid, 1984; Matsuda, 2001; Wilkening, 1981) have explored the relationships among distance, speed, and duration which require participants to judge duration based upon the interaction between distance and speed. In contrast to Levin who alleged that Piaget had too many confounding variables, Wilkening (1981) says that Piaget's isolation of one variable for study made the task less realistic since in everyday experience one must deal with distance, speed, and duration simultaneously. The overall design of these studies is similar. A scenario is set up in which animals travel a certain distance in response to a barking dog (Acredolo et al., 1984; Wilkening, 1981) or trains of different colours travel along a track (Matsuda, 2001). In Wilkening (1981) and Matsuda (2001), participants only had to deal with the movement of one animal or train at a time while in Acredolo et al. (1984) children had to compare distance, duration, and speed for two animals. In each study one of the three variables (distance, duration, or speed) is the dependent variable and relationships between the other two are varied. A duration example from Wilkening (1981) should clarify. Participants were shown a picture of a turtle, guinea pig, or cat that was placed 70 cm from a picture of a dog. Children were asked to press a button that would produce a recording of the dog barking. They were to keep pressing the button long enough to indicate the amount of time the animal would have had to run in response to the barking to travel that distance. It was expected that they would press the button longer for the turtle than the cat if they understood that a cat could travel at a faster rate and would therefore traverse the 70 cm in less time than the turtle.

Wilkening found that 5-year-olds, 10-year-olds and adults could all infer distance correctly if given speed and time. In fact, graphs of responses show that participants at all three ages knew that the difference between the distance travelled by the turtle and the cat in the task described in the previous paragraph would be greater for a longer period of time than for a shorter one. Wilkening concludes that even the 5-year-olds were employing a multiplicative strategy. That is to say, they knew that if an animal travelled twice as long at a constant speed, it would travel twice as far. The younger children did experience difficulty integrating speed and distance to infer duration. They did not clearly understand that two animals travelling at different speeds would require different amounts of time to traverse the same distance. His results suggest that the direct relationship where distance = speed x duration is easier than the inverse relationship of duration = distance ÷ speed.

Matsuda's (2001) results with Japanese children provide a cross-cultural comparison of Wilkening's data. The youngest children in her study were 4 years old, and the oldest were 11 years, 11 months. Like Wilkening her sample included adults for comparative purposes. None of the adults in her sample made errors on any task. She found that a majority of the youngest children could correctly deduce direct relationships such as determining distance if speed varied but duration was held constant or duration if distance varied but speed was held constant. Consistent with Wilkening's results, the youngest children made fewer correct responses for questions requiring the use of an inverse relationship between distance and speed to determine duration.

Unlike Wilkening, Matsuda asked participants to justify their responses and the examiner provided feedback to them. Even though many of the youngest children did well on some of the tasks they had difficulty articulating why their answer was correct, which is not surprising. At every age group, including adults, individuals referred to speed more frequently than either of the other attributes in the justifications for their responses, but there was no significant increase in the number of references to speed from 7-year-olds to adults (Matsuda, 2001, p. 469). She hypothesizes that children may be employing a "more A, more B" strategy (Stavy & Tirosh, 2000) which works for the direct relationship but produces an incorrect response for the inverse one where the correct idea is "more A, less B." In the former case, greater speed results in greater distance. In the second case, greater speed results in a shorter duration. Further, for the direct relationship the correct answer could be obtained by attending to only two of the variables without taking the third variable into account, a strategy that Matsuda describes as a two-by-two relation. This strategy works for inferring duration if distance varies and speed is held constant. At the same rate of speed it will take longer to run a greater distance than a shorter one.

In a follow-up longitudinal study in which she interviewed children yearly from kindergarten through sixth grade, Matusda (2001) reports that even though some 11-year-olds referred to all three variables in their justifications for responses they did not do so consistently. Rather, they focused only on the two-by-two relation and ignored the third variable. She interprets her findings thus,

> *They used the coordination of two-by-two relations in the simple task to reduce processing load and answered the first problem in the complex task without switching over to the duration-distance-speed system.* (p. 478)

Acredolo et al. (1984) told first through fifth grade (6-11 years) children a story about a farmer who has a dog to guard his cabbages from skunks and rabbits. They were given two of the three bits of information (distance, duration, or speed) and had to determine the third. Consistent with the findings reported above, problems involving the inverse relationship between speed and distance proved more difficult than those involving a direct relationship. Moreover, in line with others (Piaget, 1969; Wilkening, 1981) distance was deemed to be such an important feature that children attended to it and ignored the inverse relationship between speed and duration. In line with Matsuda (2001), these authors suggest that the direct relationship could be obtained by only considering the relationship between distance and speed or distance and duration while ignoring the third dimension. In contrast the inverse relationship between duration and speed requires that all three variables be accounted for. The coordination of all three ideas requires processing capabilities that younger children may not possess. This could also explain why some researchers (Berndt & Wood, 1974; Levin, 1982) found that younger children *could* solve duration problems correctly if some of the interfering cues were removed. In some situations one of the dimensions can be ignored and the correct answer will be achieved but sometimes that is not the case.

A more recent study (Casasanto & Boroditsky, 2008) explored the relationship between space and time for duration. This is a follow-up to Boroditsky's earlier work (section 2.2.2) dealing with succession in which she concluded that spatial thinking had a greater influence on temporal judgments than the other way round. The authors report on six experiments designed to test whether there is a similar asymmetric relationship between space and time for duration. All subjects were students at the Massachusetts Institute of Technology. In the first four experiments, students watched a line "grow" horizontally across a computer screen and then disappear when it reached its maximum length. Durations ranged from one to five seconds. Conditions were varied across experiments. In some cases, they were asked to indicate how long the line was at its maximum length (distance), and other times they were asked to state how long the line appeared on the screen (duration). Sometimes students knew ahead of time if they would be asked about duration or distance; other times they didn't. In one version there were delay periods before and after the line grew, and in another a tone played concurrently with the line growing.

In all experiments, the effect of spatial information on judgments of duration was highly significant at the $p < 0.001$ level. Longer lines were judged to have longer durations than shorter ones. The effect of duration information on distance was not significant. Students did not equate shorter durations with shorter lines to the extent they viewed short lines as indicative of short durations. Overall, students' judgments about duration were quite accurate although errors followed a particular pattern. Shorter lines were judged to have taken less time and longer lines were judged to have taken longer time than they actually had to "grow" across the screen. This finding is quite similar to the forward and backward telescoping found by Janssen, Chessa, and Murre (2006) for succession that was reported in section 2.2.2. In both cases, shorter time periods were compressed and longer ones were expanded.

Casasanto and Boroditsky (2008) conducted two follow-up experiments to test whether the use of a moving line affected results. In one experiment, students watched a dot move across the screen. Again, results were the same as in the first four experiments. In a final experiment, students viewed stationary lines on a screen that were visible for a time and then removed from the screen. Participants were asked to either estimate the length of the line or to judge how long it had been visible on the screen. Once again, the spatial length of the line had a far greater impact on judgments of duration than the amount of time the line was visible had on judgments of length. An ANOVA indicated that the size of the asymmetry was similar across the six experiments.

The authors very nicely sum up their own results:

> *Piaget concluded that children could not reliably distinguish the spatial and temporal components of events until about age nine. Like many contemporary results in cognitive science, our findings suggest that Piaget was right about the phenomenon he observed, but wrong about the age at which children resolve their confusion: apparently MIT undergraduates cannot reliably distinguish the spatial and temporal components of their experience, either (p. 588).*

There are several possible reasons for why space (distance) and time (duration) are often confused. Levin (1992) says it is because they both accumulate.

Friedman alleges that adults rely on both distance and image-based processes for understanding time (1992). Both are spatial ways of comprehending time. Liberman and Trope (2008) have taken the notion one step farther and maintain that both spatial and temporal distances are at root psychological distances. Thus, spatial or temporal distances that are closer to the psychological reference point are magnified, while distances farther from the reference point are compressed.

2.2.4 Montangero and diachronic thought

The work of Jacques Montangero (1996; Pons & Montangero, 1999) is discussed separately since his ideas were used by two major researchers in the area of deep time conceptions (Dodick & Orion, 2003a, 2003b). Montangero has adopted a different framework than that of the conventional time research described in previous sections. As such, his work must be juxtaposed against what has been previously discussed, particularly the work of Piaget.

In contrast to Piaget and other researchers, Montangero (1996) conceived of temporal thinking as a separate type of reasoning related to but distinct from operational thought. He introduced the term “diachronic thinking” or the “diachronic approach” to explain how one understands present events within the context of the larger process of which they are a part. For him, the ability to employ a diachronic approach necessitates the ability to conceive of a series of actions as representing one event that occurs over a period of time. He agrees with Piaget that children progress in their thinking about time but posits the shift to be a bit later than Piaget at around age ten with a reasonably well-formed conception of time by ages eleven or twelve in terms of children’s understandings of biological and physical processes, which is consistent with what others have found (e.g., Levin, 1982) Yet, it is his contention that diachronic thought, while related to operational thought, is a distinct reasoning ability (Pons & Montangero, 1999, p. 193) that must be addressed.

He identifies four schemes that determine diachronic thought (Montangero, 1996, p. 166-169).

1. Transformation: including a quantitative change, i.e., either an increase or decrease in size or a qualitative change (e.g., the shape of a tree)
2. Temporal organization: involving the capacity to link stages in an evolving process in a time-ordered sequence and taking into account that some processes are linear while others may be cyclical
3. Interstage linkage: including both cause and effect relationships and those prerequisite actions that are more a matter of convention than necessity. For example, while many narrative stories begin with a description of the setting, it is equally possible to begin with an event.
4. Dynamic synthesis: indicating the ability to determine that a series of stages effectively represents a single event over time

As evidence for his contention of a separate reasoning ability, he and Pons compared eight- to twelve-year-old children's thinking on diachronic tasks with their ability to solve problems necessitating operational thought (Pons & Montangero, 1999). There were three diachronic tasks. In the first, children were asked to depict the life of a tree with a series of pictures. While there was no real difference in the number of pictures drawn by children at different ages, younger children tended to draw a series of pictures that depicted quantitative change only. They drew a series of trees that all looked similar to one another but increased in size. In contrast, older children drew a series of pictures that depicted both quantitative and qualitative changes. They began with a bud, then a small sapling, a small tree, and finally a larger fully formed tree. A similar experimental design with 9-, 12-, and 15-year-olds is described in Montangero (1996). In this instance children were also asked how long the tree's growth process takes. Twelve-year-olds thought that the tree grew over a longer duration than younger children did. They could conceive of a growth process requiring 100 years while the 9-year-olds mentioned times closer to their own lifetimes, such as ten years. Furthermore, younger children tended to view the time period between successive stages in the growth process as constant. In other words, they thought the rate of change remained the same throughout the life of the tree.

In the second task reported in Pons and Montangero (1999), children were shown a series of six pictures illustrating a person's day trip to the beach. The first picture showed the person arriving in the morning as the sun was just coming up over the horizon. The fifth picture showed the person leaving the beach as the sun was setting. The final picture showed the individual in bed with the moon and stars visible from a window. Children were told to draw the rays of the sun in each picture and to draw more rays the hotter they thought the sun would be. Finally, they were given pink, red, brown, and white crayons and asked to colour the person's skin in each of the pictures. They were instructed that they did not need to use all the crayons and were told that in the first picture (arrival at the beach) the person's skin was white (a culturally sensitive statement). Finally children were directed to the pictures of the day trip to the beach and asked, “If you take all these pictures together, what do they represent?” (p. 195).

Younger children saw a direct relationship between the heat of the day and the redness of the person’s skin. That is, they coloured the skin its reddest during the hottest part of the day and progressively less red as the day wore on with the skin returning to white in the final picture. Older children were able to divorce the effect from the cause and realize that if the person was sunburned, the skin would remain red throughout the entire day and evening.

Another set of tasks reported in the same study was designed to determine children’s operational thought. The first task was a probability task in which children were shown collections of black and white tokens in different ratios and asked in which cases someone with eyes closed would be more likely, less likely, or equally likely to pick a white token. In a spatial reasoning task children were shown a rotating cylinder with a pen attached to a rail above the cylinder that could be moved backward and forward along the cylinder. The experimenter described specific motions of the cylinder and pen and children were asked to draw the resulting pictures that would be produced. Responses on diachronic and operational tasks were sorted into three levels, the first indicating a reliance on only one variable (Level I), the second a transitional stage in which there is minimal attention to two variables (Level II), and the third in which multiple variables are accounted for (Level III). Pons & Montangero report a high degree of correlation among the three diachronic tasks and between the two operational tasks, but a weaker correlation across diachronic and operational tasks. They conclude, “The results show that it [diachronic thought] is actually a specific form of reasoning in spite of its relationship with operatory thought" (Pons & Montangero, 1999, p. 196).

Montangero (1996) acknowledges possible objections to his assertion that diachronic thinking is a separate reasoning ability. While he argues for this separate capacity, it is also possible to view a concept of time as one manifestation of operational thought. In describing results for both diachronic and operational tasks, he and Pons report that across similar tasks (diachronic or operational), children were likely to achieve the same level and very rarely differed by more than one level across the two types of tasks. (Someone who was at the transitional Level II on the diachronic tasks was often also at Level II on the operational tasks.) This minimal variation makes it difficult to definitively conclude that diachronic thought is a separate form of reasoning distinct from operational thought. Among those demonstrating different levels across diachronic and operatory tasks, about one-third achieved a general operational level that was higher than the diachronic while 16 percent achieved a higher diachronic level than operational. If one takes a Piagetian view of development, this could be interpreted to mean that some children are in a transitional stage. This explanation is bolstered by the fact that the authors describe age as the "common denominator of diachronic and operatory reasoning (p. 198). A wider variety of tasks would be useful to help determine whether a distinction in children's abilities across diachronic and operational tasks can be justified. For indeed, children who do not yet possess an operational understanding exhibit instability of responses (Piaget, 1969, p. 113).

Furthermore, the diachronic tasks themselves can be reinterpreted as problems of duration and succession. In the first task in which children had to draw a series of pictures to depict the life cycle of a tree, children were deemed to be at Level I if the size of the tree changed in each drawing but not the shape. Level II children were those in which the size of the tree changed in each drawing and shape changed in a minority of drawings. Level III children were those who changed both the size and shape of the tree in successive drawings. The authors appear to be making the assumption that these changes in the types of drawings must be due solely to changes in children's reasoning ability. Yet there is no evidence that children were asked to explain their drawings. Perhaps differences in drawings can be explained on the basis of improved motor skills and/or a general ability to simultaneously consider multiple aspects of change and not merely temporal ones. They may have been able to successfully order a series of drawings that depicted both qualitative and quantitative changes in a tree over time. The authors do not adequately demonstrate why other explanatory mechanisms are inadequate.

Another possible explanation is simply that older children bring to the tasks a greater number and variety of experiences with the phenomena in question. While none of the children are likely to have watched any particular tree go through a full succession of growth stages, older children will have had opportunity to see a particular tree change over time and to have seem more trees at different stages of the growth process. They also have greater experience with the qualitative and quantitative changes that are part of growth for any living organism including themselves. They are likely to have had more school science experiences with plant growth. Older children can generally be expected to possess a better concept of trees as representing a class of objects with similar characteristics. According to Friedman (1990), the fact that young children possess such a short time frame of experiences themselves means that their notions of past events are more undifferentiated.

Consider the day at the beach task described previously. In order to reach Level III children must be able to distinguish between the duration of the cause (the height of the sun in the sky representing the passage of time) and the duration of the effect (how long the skin stays red). Younger children conceived of a co-variance between the two. As the sun gets higher in the sky, the skin gets redder, and as the sun sets, the skin returns to its original white colour. In reference to an earlier, similar study, Montangero ascribed this phenomenon to children's confusion between the progression of time and the process (1996, p. 66) and cites this as evidence for diachronic thought as a construct. However, it could also be inferred that younger children have had fewer sunburns than older children and therefore are not reasoning based upon a lack of diachronic thought as much as a lack of experience. Thus, children's responses may not be indicative of a poor concept of duration. Rather, they simply don't know that the effects of some actions outlast the actions themselves. They are equating duration (the amount of time the skin is red) with the action itself.

Additional questions exist as to whether or not specific responses to experimental tasks indicate faulty understanding or merely a child who relates only the main ideas of a story, as some children do. Montangero's 1996 work provides an example. Children were shown a picture of the familiar storybook characters Babar and Celeste on pieces of ice. They were provided with additional sheets of paper containing a picture of Babar and Celeste and were instructed to draw what happened before and after the original picture and then comment on their drawings. The following is an example of one child's description of his drawings and is described by Montangero as a scenario two response (out of a possible four):

- *Babar and Celeste are ice skating.*
- *The sun comes out and starts to melt the ice.*
- *The sun gets hotter and hotter and makes the ice melt more.*
- *The sun gets so hot that it melts all the ice and Babar and Celeste fall in the water.* (Montangero, 1996, p. 61)

This is contrasted with a scenario three response which is deemed to evince a higher level of diachronic thought.

- *Babar and Celeste are ice skating, they are making marks on the ice.*
- *They go on skating; they make more and more marks. The ice which is too thin is beginning to crack.*
- *The more they skate, the more the ice cracks. Babar and Celeste are left on little bits of ice.*
- *They fall in the water.* (p. 61)

Scenario three is more elaborate than scenario two, yet it is not obvious why it is determined to be of higher conceptual quality. Montangero states that it is because it demonstrates a dissociation of the time of the cause from the time of the effect. Without additional information about the children's responses, this is a difficult claim to accept. One would like to ask both of these children to explain their answers more fully. It could be that there are no real differences in their reasoning abilities. Instead, the difference in their answers may simply indicate that one child spontaneously provides more detailed responses than the other. Montangero has not clearly defined his criteria for the reader. This makes it difficult to judge the validity of his evidence, and consequently, his theory.

Montangero's findings could also be seen as similar to Piaget's results with children who said that it took more time to draw a large number of lines quickly than fewer lines slowly even though the time interval for both was fifteen seconds (Piaget, 1969). The ability to realize that fifteen seconds is a standard interval regardless of the amount of work accomplished is a characteristic of an operational understanding of time. Therefore, it seems unnecessary to invoke a separate reasoning ability known as diachronic thinking to explain the progression of children's thinking and reasoning about time.

2.2.5 Summary of the research on conventional time

Researchers from Piaget (1969) forward have shown the importance of an understanding of succession (the temporal order of events) and duration (the length of time of events or processes) to an understanding of conventional time. A concept of duration requires the ability to account for differences in distance and/or speed. Some, like Piaget have isolated the two ideas. Others (Acredolo et al., 1984; Matsuda, 2001; Wilkening, 1981) have looked at the relationships among duration, distance, and speed. While very young children possess rudimentary concepts of both succession and duration, the age at which that can be consistently applied for many time scales is clearly later than Piaget thought. For reasons that are not completely obvious, even adults do not always display the operational understanding of time that would be expected. Nonetheless, the body of research on conventional time supports Piaget's contentions that the notions of succession, duration, and the ability to mentally move forwards and backwards to reconstruct events are sound. Individuals at all ages appear to find the direct relationship speed x duration = distance easier than the inverse relationship distance ÷ speed = duration.

There are strong suggestions that people map time spatially and that both spatial and temporal distances represent psychological distance. Events in the past or future are viewed as temporally distant from oneself in the present or some other event that is fixed in time. There is a telescoping of events with those in the more distant past being judged as being closer to the present and to each other. This is the same phenomenon that occurs spatially as one drives down the road and sees two large hills in the distance. The hills appear quite close to one another when one is at a distance from both. It is only as one gets closer to the first hill, and the reference point changes that one can see the two are actually quite far apart. In the same way, from the vantage point of the present, events in the past are viewed as closer in time to one another than they really are.

No compelling reason can be found to argue for a separate reasoning ability known as diachronic thinking. Minimally, the strong relationship between how people reason spatially and temporally argues that temporal reasoning is not distinct from more general reasoning. If one alleges a separate temporal reasoning ability, then it would also be necessary to propose that spatial and temporal reasoning differ from each other in significant ways. At present, the data does not support such a claim. The validity of diachronic thinking is also called into question by the fact that many tasks deemed to provide evidence for its existence can just as easily be interpreted as succession and duration tasks. Thus, the research from Piaget to the present indicates that an understanding of conventional time is built upon the twin understandings of succession and duration. In situations where there is a direct relationship between the variables involved, making a judgment about succession or duration is fairly straightforward. When an individual must determine how to reconcile competing pieces of information (an indirect relationship) to judge succession or duration, the task is more complicated. In section 2.5, we will explore how these ideas relate to deep time. First, we consider research into how people understand large numbers.

2.3 Large numbers

> *For all practical purposes, an $800 billion stimulus package is as opaque as a 703,000-hectoshekel package; we have no real grasp of what it means. Big numbers fuzz our brains, and that is just as true in business as it is in public policy. Speaking in "millions" and "billions" is like your second year of Spanish: You've memorized the vocabulary, but it's hard to think in the language.* (D. Heath & C. Heath, 2009, p. 59)

We are surrounded by large numbers in many areas of life. Yet, intuitively we aren't sure we really understand them. Heath and Heath's use of the phrase "fuzz our brains" captures that view. Why do many humans possess such poor number sense for large numbers? To answer that question it is important to briefly describe what constitutes number sense for numbers considered "smaller."

Among other things, number sense involves the ability to determine the relative magnitude of numbers and understand how two numbers relate to one another. Elementary (primary) teachers spend a considerable amount of time helping children explore how numbers are related to one another. Good number sense requires a basic understanding of the decimal number system in which multiples of ten function as benchmarks to which other numbers can be related. As children get older they must also develop an understanding of the proportional relationships between numbers. Proportional reasoning enables someone to not only determine that 12 is three times greater than four, but also to move multiplicatively through orders of magnitude in the base-ten system. This section reviews literature on how students relate numbers of various magnitudes to one another and how they understand proportional relationships between numbers. A number of studies have been conducted primarily with children and illustrate how younger pupils understand relationships amongst numbers. Research into adults' understanding of number extends the findings of the studies with children.

2.3.1 An understanding of number and relationships among numbers

One strand of research into how people understand number deals with how individuals map numbers onto space, or a number line. The majority do not deal with numbers in the range involved in deep time, but they provide useful parallels. A number of the studies (Booth & Siegler, 2006; Dehaene, Izard, Spelke, & Pica, 2008; Kadosh, Tzelgov, & Henik, 2008; Laski & Siegler, 2007; Petitto, 1990; Siegler & Opfer, 2003) employed a similar research design that varied slightly across studies. Children were shown a number line on which the endpoints, either 0 and 100 or 0 and 1,000 were marked. They were subsequently asked to place specific numbers on the number line in their appropriate place. Kindergartners used what Dehaene, et al., (2008) terms a compressed logarithmic scale. This does not mean that young children employ a true logarithmic scale in the sense of a slide rule which would represent a sophisticated understanding, to be sure. Rather it is more akin to the example of a decreasing interval scale as illustrated in Figure 2.1. This figure does not represent the scale used by any individual child. It does, however, nicely illustrate what is meant in the literature by a "compressed logarithmic scale." The differences between 0 and 10 or 10 and 20 are exaggerated while differences between numbers closer to 100 (or 1,000) are diminished.

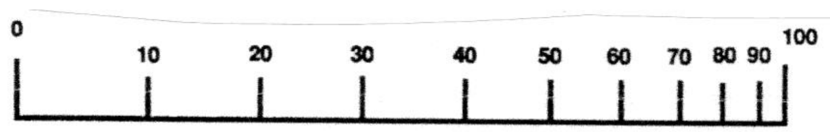


Figure 2.1 Decreasing interval scale, (Petitto, 1990, p. 73)

Weber's law, which says humans perceive differences between smaller numbers more easily than between larger ones, is often invoked to explain the reason for this logarithmic mapping. Larger numbers must be a greater proportional distance apart to be discriminable, (Dehaene et al., 2008, p. 1219). I am well aware that five and seven differ by the same amount as 55 and 57. I can readily distinguish the difference between five and seven sweets lying on the table. Yet, if I see two piles on the table, one containing 55 and the other 57 sweets, I may not be able to distinguish that there is a difference between the piles.

From a perceptual standpoint why might this be? Consider the role of the reference point 0, or the origin. If I am standing on a number line at 0, a person standing at 5 is easily perceived as closer than someone standing at 7. However, two people standing at 55 and 57 will likely appear much closer together than the people at 5 and 7. If we think of the distance from the origin to the larger of the two numbers in the pair as our whole distance, the reason for this may become clearer. The person standing at the 7 represents the entire distance in the first pair of numbers. The person standing at the 5 is at 5/7 of the entire distance. In contrast, if the difference between 0 and 57 is taken as the whole distance when comparing our second pair of numbers, then 55 is 55/57 of the entire distance. In other words, from the frame of reference of 0 (the origin), 55 is closer to the whole distance of 57 than 5 is to the whole distance of 7.

There is a striking parallel between a logarithmic mapping of numbers and the temporal compression of events described in section 2.2. Thus, it could be argued that numbers like time are mapped spatially. The analogy of the two hills in the distance to temporal distortion was described in section 2.2.5, but could be equally applied to a logarithmic mapping of numbers. The reason why the two hills are judged to be closer than they actually are can be explained on the same basis as the numeric example in the previous paragraph.   The issue is my current frame of reference.

Children's thinking appears to shift from the logarithmic to a more linear representation somewhere between second and fourth grade for the 0-100 number line, depending upon the study. That means that children are then able to map equal intervals, and the distance between 7 and 18 is seen as the same as the distance between 54 and 65. However, these children still produce a more logarithmic scale for a number line from 0-1,000. By sixth grade, most children are able to place numbers on the 0-1,000 number line in a linear fashion. Increasing experience with larger numbers allows children to expand their frame of reference. Size of a number is relative to other numbers to which it is being compared. One hundred is a large number if I am comparing it to 10. It is not so large if I am comparing it to 1,000.

The issue of relative size shows up in another way in this body of research. The scale on which a number is placed appears to influence how children represent it. There is some evidence that children represent the same number differently depending upon which number line was used. Using the experimental design described previously, Siegler and Opfer (2003) asked children to place a subset of seven numbers on both the 0-100 and 0-1,000 scales. They found that some children were able to represent a particular number linearly on the smaller number line but represented that same number logarithmically on the larger one. For example, a child might site 50 midway between 0 and 100 on the smaller scale, but place the same number much farther from 0 than would be appropriate on the 0-1,000 scale. These results seem to hold cross-culturally (Dehaene et al., 2008) although the age at which the transition to a more linear scale occurs is not identical. Siegler and Opfer (2003) interpret their results as evidence for the use of a landmark proportionality model to map numbers on the number line. Individuals divide a number line along specific reference points that are then used as landmarks to place other numbers on the line. The base-ten system lends itself to just such a strategy. We often refer to reference points (usually numbers ending in 0) as landmarks or “nice numbers” when speaking with children and encourage them to use those “nice numbers” in flexible computational strategies.

Adults also exhibit logarithmic mapping in certain situations. Dehaene et al. (2008) extended the experimental design with number lines described above to an Amazonian culture. Their study included both children and adults. Their findings were highly consistent with earlier studies with children. In fact, adults as well as children mapped numbers logarithmically. It was not unusual for adults in their study to perform similarly to young children in the studies described in the last few paragraphs. While age did not correlate well with performance, educational level did. More highly educated participants were more likely to map numbers in a linear fashion than those who were less well-educated. The authors offer two possible hypotheses to account for these results. Students have experiences with standard units of measurement in school. These standard units serve as benchmarks or constants that permit two people to measure the same object and get the same result. Dehaene et al. (2008) also allege that experience with addition and subtraction fosters the move from a logarithmic to linear concept of numbers. This is related to the first idea because it involves the understanding that consecutive whole numbers are related to one another by the standard unit +1 or -1. I would add that experience with multiplication and division is also important. This is particularly true as the scale is extended to larger numbers. A study that employed a slightly different experimental design than what has been previously described illustrates why that is so.

Confrey (1991) reports on a case-study interview with a female undergraduate. The purpose of her paper is to illustrate how researchers use problem solving tasks to infer student conceptions. Thus, Confrey makes no claims that this young woman's conceptions are normative. That said the type of reasoning in which the student (Suzanne) engaged is useful for the present discussion.

Suzanne participated in several interviews. Two sessions were designed to probe her understanding of scientific notation—specifically how exponents represent orders of magnitude. She was asked to place a series of events from the Big Bang to the Renaissance on a timeline and could use as much paper as she needed to complete the task. Suzanne was provided with the age of each of the events, thus neither her geoscience nor her historical content knowledge was being assessed. Still, the task may have been complicated by the fact that numbers less than one million were written in standard notation while those greater than one million were written in scientific notation. (Standard notation in the U.S. means writing a number in the form 2,000. That same number would be written as 2×10^3 in scientific notation.) Suzanne expressed confusion about the fact that some numbers were written one way while others were written in a different manner. At various points she tried converting all numbers in standard notation to scientific notation or all those in scientific notation to standard notation, but was unsure if she was correct. Early in the interviews she tried to place 1.5×10^{10} midway between 1.0×10^{10} and 1.0×10^{11}. However, in reality that number is midway between 1.0×10^{10} and 2.0×10^{10}. Thus, even on a logarithmic scale her solution was incorrect.

Suzanne initially attempted to place all events on one sheet of paper even though she could have used multiple sheets. She divided up the page into intervals so she could do so but lost the relative, proportional scale as a result. Although she stated that the value of numbers between her intervals increased by a power of ten (logarithmically), she struggled to interpret distances logarithmically. Instead, she interpreted them linearly, at least at first. She concluded at one point that dinosaurs were around for a shorter amount of time than humans because the space between their appearance and extinction was shorter than the space between the appearance of humans and the present. After probing, she was able to reason her way to the correct answer by switching the numbers from scientific to standard notation although the conversion proved difficult. By the conclusion of the second session she had moved to a more linear scale although it was still somewhat hybrid. Perhaps the use of scientific notation for numbers greater than one million predisposed Suzanne to use a logarithmic scale. While that cannot be ruled out, other results (Dehaene et al., 2008; Holyoak & Mah, 1982) suggest that a compressed logarithmic scale is not atypical for adults when dealing with larger numbers.

This compression of larger numbers and expansion of smaller ones appears to hold true in a variety of situations. It does not seem to matter whether the larger number is on the right or the left, thus it is not merely an artefact of a left to right orientation. Even the larger number can be infinitesimally small from a deep time context. Adults, 5-year-olds, and 7-year-olds were asked to bisect a line that had two quantities on either end (deHevia & Spelke, 2009). One version of the task used the Arabic numerals 2 and 9 while another used two dots and nine dots. Order of presentation and the side on which the larger quantity was presented were varied. Adults bisected the line toward the larger quantity irrespective of whether it was the Arabic numeral or dots. Seven-year-olds did the same for dots but the pattern was less consistent for numerals, perhaps because the numerals held less meaning for them as representing specific quantities. Five-year-olds, who were only tested with dots, also bisected the line closer to the larger quantity. These results are consistent with those described previously. Clearly, individuals could solve the problem in one of two ways. First, they could actually calculate how many numerical units there are between two and nine, determine the midpoint, and bisect the line appropriately. Second, they could simply bisect the line spatially without any reference to the quantities involved. If they were employing the latter strategy exclusively they would be expected to either bisect accurately or to show no real preference to bisect toward one number rather than the other. Neither of those possibilities is what happened. The fact that participants at all three ages tested bisected the line closer to the large quantity lends support to the contention that people appear to compress the difference between larger numbers and expand the difference between smaller ones on a number line. That holds true even when the numbers in question are quite small or when those quantities are represented by objects like dots.

The use of reference points or a “landmark proportionality model” (Siegler & Opfer, 2003) appears to be hold true in situations not involving number lines. Izard and Dehaene (2008) describe a series of experiments with French young adults. Participants were shown arrays of 9, 30, or 100 dots that were scattered in different arrangements on a computer screen. They were then asked to estimate the number of dots in each array. All participants consistently underestimated the number of dots in the arrays, but the degree of underestimation varied from individual to individual. Further, estimates across individuals varied more for 30 dots than for 9 and more for 100 than for 30. In a follow-up experiment designed to determine whether providing individuals with a specific referent would influence their estimation judgments, subjects were first shown an array and told it contained 30 dots. Sometimes the original array did contain 30 dots. Other times, even though subjects were told there were 30 dots there were actually either 25 or 39. Next, participants were shown dot arrays on a computer screen identical to what was described for the first experiment. Subjects consistently adjusted their estimations for all dot arrays (no matter if they contained 9, 30, or 100 dots) to the referent even if the referent was wrong. Those who were shown an array with 25 dots but were told it contained 30, underestimated the number of dots in subsequent arrays. The reverse was true for those who were told the referent array contained 39 dots.

This reliance on a reference point does not only seem to occur with numbers or dots. Holyoak and Mah (1982) investigated how people use reference locations to determine how much closer or farther one of two cities was from a reference point, (the Atlantic or the Pacific Ocean) than the other city. When the reference point was the Pacific Ocean, university undergraduates rated the distance between two western cities as greater than they rated their distance from each other when the reference point was the Atlantic Ocean. Further, when asked which of two cities was closer to the Pacific Ocean, response times for western cities were shorter than response times for cities in the middle or eastern part of the U.S. When the Atlantic Ocean was the reference point, the converse was true. The differences between cities that were closer to the reference point were magnified while distances between those that were farther away were compressed.

These results could be an artefact of the experimental design but that design might also explain why the phenomenon occurs. When subjects in Holyoak and Mah's study compared two cities they had to choose a number between one and nine to indicate how much closer one city was to the reference point than the other. A rating of nine meant that one city was maximally closer to the reference point than the other. Unfortunately, no specific examples of individual responses are provided to see how this played out in the experiment. However, their results may indicate how people use reference points to create units and explain why there is this expansion and compression of distance in relation to the reference point.

For ease of discussion, assume a hypothetical situation in which the Pacific and Atlantic Oceans are 1,000 kilometres apart (not true in North America). Four cities which I will call R, S, T, and U are situated roughly on an east-west line between the oceans, each 200 kilometres from the next closest city. If I am asked to determine how much closer R is to the Pacific Ocean than S, the Pacific is clearly my reference point. I reason that when I have reached R, I must still travel twice as far to reach S since R is 200 kilometres from the Pacific Ocean while S is 400. I may not even think in terms of actual distances, but rather in terms of time. If it takes me a certain amount of time to reach R, I will need to travel twice as long to get to S. Now reverse the scenario. How much closer is S to the Atlantic Ocean than R? S is 600 kilometres from the Atlantic while R is 800. With the Atlantic Ocean as my reference point, I can now imagine the distance to S or more probably the amount of time it will take to get from the Atlantic Ocean to S.   When I reach S, I must only travel a little farther (or longer) to R than the distance I have already gone, not twice as far (or long) as when the reference point was the Pacific Ocean.

Earlier in this section, I hypothesised that people may view 7 and 9 as being farther apart than 55 and 57 for a similar reason. In the first case seven is 7/9 of the distance to nine (if 0 is my reference point), while 55 is 55/57 of the distance to 57 (again, with a reference point of 0). Thus, from the vantage point of 0, the distance between seven and nine is greater than the distance between 55 and 57. The examples with pure numbers and distances are analogous but not identical. Consider the numerical example first. Here, my prior experience with numbers plays a crucial role in my ability to conceive of those two numbers as equidistant apart. To place all four numbers on a linear scale, I must essentially be able to mentally step outside of the scale and see the entire scale at once. It's reasonable that younger children who have had limited experience with large numbers would find that a difficult task.

For distances, as in the Holyoak and Mah (1982) study, my experience with numbers, i.e., knowing that 400 is twice as large as 200, is not the only factor that influences how I perceive the task. A second component of my understanding is my experience traversing spatial distances which I often interpret as temporal distances. If it takes me two hours to get to R from the Pacific Ocean, then it will take me four hours, or twice as long to get to S. In relative terms, I am still a long way from S. Conversely, if it takes me six hours to reach S from the Atlantic Ocean, I will only have to travel a short time to reach R. I am almost there. Thus, spatial and temporal understanding are related as we saw in section 2.2. From the studies reviewed in the current section, it appears as if people perceive numbers in analogous ways to how they perceive time and space. Again though, in order to accurately complete Holyoak and Mah's task by placing the distances on a linear scale, I cannot situate myself anywhere on the scale. I must step outside the scale and, in essence, view the five cities and the two oceans from space. Then, I can place the distances between cities and oceans on a linear scale. The third component that comes into play on this task will be discussed more fully in section 2.4 but is briefly mentioned here. I must also know something about geography and the five cities in question. There is an interaction between my specific geographical content knowledge and my knowledge of numbers. Each of these facets—my knowledge of numbers and my knowledge of geography—provide me with referents that I can use to evaluate my placement of distances on the scale.

The reader may ask whether it even matters if people map large numbers using a logarithmic rather than a linear scale. One aspect of number sense is the ability to distinguish amongst numbers of different magnitudes and to be able to use that distinction to judge the plausibility of answers to calculations. If a child has good number sense for two digit numbers, it will be clear that the answer to 46 + 67 could not be 1,013. Anyone who has worked with primary school children will know that not all children realise this answer is unreasonable! Laski and Siegler (2007) argue that linear mapping of numbers is critical to this notion of number sense since it permits the discrimination of the size of numbers throughout a large range. In contrast, logarithmic mapping means that larger numbers are not distinguished from one another but are simply viewed as "all those big numbers," (p. 1740). In that compressed logarithmic scale 1 million and 1 billion are either not distinguished from one another or the difference between them is seen as less significant than it actually is.

As students age and have more experience with larger numbers they are more able to deal with larger numbers on a linear scale. One reason is because they possess a greater variety of larger, meaningful reference points. The fact that many adults don't have those referents for very large numbers means it is likely that a logarithmic representation persists into adulthood for those quantities.

> *It may be the case that people never automatically use linear representations of numerical magnitude for all types of numbers. Instead when operating in unfamiliar numerical ranges, even adults may need to override the impulse to use the logarithmic representation…Thus, adults, as well as elementary school children may rely on logarithmic representations in unfamiliar numerical ranges. The frequent confusion of millions, billions, and trillions in news magazines and political discussions lends plausibility to this prediction.* (Booth & Siegler, 2006, p. 200; see also Dehaene et al., 2008)

In addition to a linear scale and meaningful reference points, students need facility with proportional reasoning to be able to deal with very large numbers. Tourniaire and Pulos (1985) reviewed some of the literature on proportional reasoning. Proportional relationships are multiplicative. Tourniaire and Pulos found that a common error in responses across studies was the use of an additive problem solving strategy rather than a multiplicative one. Suppose the task is to write a ratio that is equivalent to 5/7. Instead of using a multiplicative strategy in which both the five and the seven are multiplied by two to yield 10/14, the student uses an additive strategy producing 7/9 (an error that is not uncommon amongst children). The size of the numbers appears to affect task difficulty for proportional reasoning as does the familiarity with the context in which the problem sits. Thus, one could expect students to have greater difficulty dealing with proportional relationships between numbers in the millions and billions than for smaller quantities. I have already alluded to the role familiarity with the context might play in my discussion of why students had difficulty with the Holyoak and Mah (1982) task. This point will be developed more fully in section 2.4 on subject matter knowledge, so I will say nothing more about it here. I now turn to a discussion of the role of a unit to an understanding of number.

2.3.2 The role of a unit

Facility with large numbers requires the use of units that are not part of daily experience. The ability to think in terms of a unit allows an individual to reinterpret a situation in light of that new unit (Lamon, 1994). This happens when younger primary school children are able to conceive of ten as not simply a collection of ten individuals but as a collective unit, so that the child can reason about one ten, two tens, etc. As mentioned in the last section, multiples of ten serve as reference points or landmarks around which other numbers sit. Many models used in science require students to conceive of and then work in different units. One group of researchers (Tretter, Jones, Andre, Negishi, & Minogue, 2006; Tretter, Jones, & Minogue, 2006) use an example of a class of 25 students. While one can think of a unit of a single individual in the class, one can just as easily work in the unit of the entire class.

A child must understand that a unit represents equal intervals in order to think in that new unit. This equal interval concept doesn't appear to develop for smaller numbers until mid or upper primary school. Petitto (1990) provided children with five rulers that were marked off in the following ways: equal intervals, increasing intervals, decreasing intervals, alternating intervals, or irregular intervals. Children were told they could choose a ruler to help them solve number line problems like those discussed in section 2.3.1. Of the 81 first-third graders in the study, only 12.5% consistently used the equal interval ruler. The percentage increased for third graders (42.9%), but that hardly represents a solid understanding of equal intervals by the group as a whole. The one unfortunate thing about Petitto's study is that she does not differentiate amongst children who did not consistently use the equal interval ruler. The decreasing interval ruler (see Figure 2.1) corresponds most closely to the compressed logarithmic scale that was described in the previous section. However, we do not know if children who did not use the equal interval ruler preferred the decreasing interval one to the other incorrect choices.

Some interesting research into the role of a unit deals with size. An initial cross-age study, (Tretter, Jones, and Minogue, 2006; Tretter et al., 2006) investigated conceptions of scale phenomena of fifth, seventh, and ninth graders from the same school district in North Carolina, a group of academically talented high school seniors from across the state attending a summer program, and doctoral students in science or science education. The first task was a paper-and-pencil questionnaire that was administered in a large group setting and was modelled after a task used by Roger Trend (2001b) in his research into conceptions of deep time which will be discussed later. Subjects were given a list of 26 objects such as diameter of a hydrogen atom, length of a school bus, and diameter of the Earth and asked to indicate their size from less than one nanometre to more than one billion metres. The second task was a card sort, individually administered, in which subjects sorted 31 cards with the name and picture of an object into as many size categories as they wished. Objects ranged from the distance from the Milky Way to the farthest galaxy to the size of the nucleus of an oxygen atom. After sorting, subjects were asked to explain their reasoning for their sorts and any previous experiences that helped them with the task.

Relative size rankings were more accurate than absolute for all age groups. Thus, participants could compare relative sizes of objects even if they didn't have a good sense of exactly how large either object was. Not surprisingly, there was greater variability within the rankings of the three youngest groups than within the two older ones which could be explained by less familiarity with objects at either end of the scale. For example, primary school students are unlikely to have had much exposure to microscopic objects and would be unable to reasonably judge their sizes. In fact, the authors note that some of the younger children sorted objects based upon the size of the picture on the card rather than the size of the object the picture represented. Without knowledge about the object in the picture, a child could only use the size of the picture itself to sort. Prior experience was a factor for the experts, as well. Those whose work is with small scale phenomena had difficulty with large scale distances and vice versa.

All groups appeared to use themselves as an important size referent and distorted sizes around “self.” Objects that were closer in size to a human tended to be ranked as more different from a human in size than they actually are. Objects larger than the self, such as an elephant, were deemed larger than they actually are and those that were smaller, such as a textbook, were rated smaller than they are. The opposite occurred at the extremes of scale. Here very large objects were rated as smaller than they are and very small objects were rated as larger than they are. This distortion is similar to the mapping of numbers onto a number line (e.g., Siegler & Opfer, 2003). In both cases objects or numbers that are farther from the reference point are judged to be closer to it and to each other than is correct.

Generally speaking, older groups conceptualized size categories that were more distinct than younger groups. All ages created three or four categories larger than themselves with the largest category being called “big” by all groups except ninth graders who had a category called “space.” Some included continental distances in their largest group. Several high school students made comments to the effect that while planetary distances are very different from one another in comparison to the size of a bedroom, for example, they are close. This is additional evidence of the compressed logarithmic scale noted earlier. The distance from New York City to Los Angeles is expressed in thousands of kilometres, the distance from the Earth to the Moon in hundreds of thousands of kilometres, and the distance from the Earth to the Sun in millions of kilometres, yet all were placed in the same size category by some students. In contrast, finer size distinctions were made between objects closer to human size. For all groups, accuracy of large group sizes declined smoothly while a different phenomenon occurred at small scale. Accuracy was high through the smallest visible level and then dropped precipitously after that.

Of particular importance to the present study is the experts' use of measurement units in describing how they categorized sizes. The following comments were typical of this group, "You know how people create new units to avoid having to think of large-scale units you can't understand," and "I would measure in units like parsecs or astronomical units," (Tretter et al., 2006, p. 302). These units serve as referents that permit someone to move back and forth between size categories. There were several specific ways in which experts employed what the authors call a unitizing strategy:

1. redefined units when it seemed more convenient to do so, e.g., changing one million meters to 1,000 kilometres
2. determined the piles in the card sort based upon the units used to measure objects in that pile
3. thought of an object that would represent that size and then used that as the reference point, e.g., one participant indicated using an atom as a reference point when thinking of objects measured in nanometres

When asked about their reasoning, subjects in the three younger groups often mentioned trying to think of the largest or smallest thing they knew as a referent. Some seventh and ninth graders reported trying to calculate sizes, particularly for large scale phenomena. The seniors evinced a transition to employing a unitizing perspective, for example, considering the diameter of the Earth and then comparing distances to that referent. The authors conclude, "From these data, it may be surmised that, to effectively engage at very large or small scales, it is necessary to mentally transport oneself to a new world at that scale," (Tretter, Jones, and Montague, 2006, p. 1081).

I would add that this is only possible if the numbers at that scale have some meaning for the learner. A confounding variable for U.S. students is a lack of familiarity with the metric system. Even though the metric system is taught in U.S. schools, it is taught alongside the U.S. customary system, with which children have extensive everyday experience. The authors say they attempted to control for lack of familiarity with the metric system by showing 5th and 7th grade students a metre stick, explaining the categories to children, and demonstrating them on a metre stick before students answered the first four questions on the Scale Anchoring Objects (SAO) task which was described above. For example, the researcher told children a decimetre is 1/10 of a metre, and pointed that distance out to them on the metre stick. The metre stick was available for the older groups of students but researchers did not go through the process of showing them a decimetre as they had with the younger pupils. It is questionable how much benefit 5th and 7th graders were able to make from such a brief tutorial. It is also likely that at least some U.S. 9th graders were unsure that a decimetre is 1/10 of a metre and would not know how to indicate that distance on a metre stick.

My contention that a unit must represent a meaningful referent for a learner is corroborated by another study. Jones, Tretter, Taylor, and Oppewal (2008) demonstrated how familiarity with standard mathematical units is important for accurately estimating size. Fifty undergraduate and postgraduate trainee teachers and 60 experienced teachers enrolled in a masters' program in science participated in their study. The authors used the same instruments described in the preceding paragraphs (Tretter et al., 2006a; Tretter et al., 2006b). Consistent with other results, both experienced and trainee teachers were able to more accurately estimate sizes at the human scale and progressively less so as they dealt with objects toward the extremes of scale. They demonstrated greater accuracy when using standard metric units, but less accuracy when using a non-standard unit (body length). Experienced teachers were more accurate at the extremes of scale when using metric units than the trainee teachers. There were no real differences between groups at the extremes of scale with non-standard units. These results lend support to the claim that familiar units act as meaningful referents to place objects.

Several of the aforementioned authors conducted a similar study with 17 visually impaired students in the U.S. (Jones, Taylor, & Broadwell, 2009). They found an interesting difference between this sample and the studies with sighted individuals already reported. While the overall trends are similar, visually impaired students were more accurate when estimating sizes at the extremes of scale than were sighted students. The authors proffer several possible explanations. One is that since neither very small nor very large scales can be experienced kinaesthetically in the way sizes closer to the human body can, students normally learn about them through auditory or visual means. Visual channels can lead to spatial distortion in the manner described in section 2.2.5 for the two hills in the distance that appear quite close together but actually are far apart. Visually impaired students are not receiving those visual cues which could distort perception. It is an interesting suggestion that appears to agree with other findings discussed in this thesis and bears further investigation.

2.3.3 Summary of the research on large numbers and the role of a unit

Large numbers or quantities appear to be perceived and mapped onto a compressed logarithmic scale that exaggerates differences between objects or events that are close to the reference point. Numerical mapping is similar to both spatial and temporal mapping. Differences between cases that are farther from the reference point are compressed. With numbers, it's often 0, but could be another benchmark, like a multiple of ten. With space it's a particular location or size referent. With time, it is often the present. In terms of number, this logarithmic mapping means there is less discrimination between larger numbers than smaller ones. Greater experience with large numbers gives them more meaning. The problem with really large numbers is that we don't have much experience with them. To many students one million and one billion are both simply very large numbers. Many older students and adults could tell you that a billion is more than a million but how much more is elusive. The proportional relationships between numbers allow for the use of units that represent fixed quantities that remain unchanged no matter what is being measured. A solid understanding of units appears to be particularly important for making sense of sizes that are outside a human's ability to directly experience.

Children in middle to late childhood move toward a more linear mapping of numbers, at least up to 1,000. Children at that same age develop clearer understandings regarding both succession and duration that enable them to take account of factors other than spatial distance to judge them. None of the issues that confuse children about numbers or time ceases to be a problem, even in adulthood. In certain circumstances, adults make both temporal and numerical errors that are more common to younger individuals.

I will discuss the role played by subject matter knowledge in concept acquisition in the next section. Yet, it has already crept into the discussion about large numbers several times. Perhaps it is experience within a field using appropriate units that is crucial in making sense of the numbers involved. Several studies discussed above have demonstrated that is true with size (Jones et al., 2008; Tretter et al., 2006a; Tretter et al., 2006b). Better knowledge of metric units and familiarity with specific units gives them meaning for a student. It's not the subject matter knowledge of any particular discipline, but it is knowledge of quantities and the proportional relationships among them. Possessing one or more mental referents for a specific unit of size or time enables one to reason about that unit. Without that, one million years is merely a very long time. We now turn to the ways in which knowledge within a particular domain influences concept acquisition.

2.4 Subject matter knowledge and concept acquisition

> *One generalization has been repeatedly supported in the research on cognition and instruction conducted during the past 2 decades. That generalization states that what individuals already know (i.e., prior knowledge) exerts a powerful influence on what they will come to know* (Alexander, Kulikowich, & Schulze, 1994, p. 314).

This statement is as true today as it was fifteen years ago when it was written, and just as true as when David P. Ausubel made a similar statement in his 1968 book on educational psychology (Ausubel, 1968). Constructivism, as a theory of knowledge acquisition, rests upon the idea that learners are not empty vessels to be filled with knowledge. Rather, they bring much to the table when confronted with new information: knowledge, dispositions, and skills that interact with new information and determine if and how the new information will be integrated into existing schema. To be sure, some of what students bring to the table aids in concept acquisition while other aspects hinder new learning. Students' current conceptual ideas are so integral to the acquisition of new ideas that some have argued they are the most important factor in determining if and to what extent new information will be acquired (Posner, et al., 1982; Scott, Asoko, & Leach, 2007).

Every section of this chapter points to how prior knowledge of something influences conceptual development. Thus far, I have described how what people understand about conventional time and large numbers affects how they are able to solve problems or complete tasks in each of those areas. This section focuses more broadly on the ways in which subject matter knowledge in a particular domain contributes to how students make sense of new information in that field. I will discuss more specifically the role each of these factors plays in a conception of deep time in section 2.5.

2.4.1 A model for the role of subject matter knowledge in concept acquisition

If we are fuzzy on just what a concept is, trying to explain the role of current knowledge in the acquisition of new concepts is a formidable task. In order to explore subject matter knowledge's role in concept acquisition it is useful to first think about the ways in which subject matter (or domain) knowledge can exist in the mind of a learner. Subject matter (domain) knowledge is multi-faceted. There are many ways to think about how it is constructed. Every learner brings knowledge from a wide variety of areas to a task. While this broader knowledge plays a role in concept acquisition, the focus of this section is on the role played by general knowledge of a domain as well as more specific knowledge of topics within that domain. That is because this thesis is concerned with the ways knowledge of geoscience impacts how students understand deep time. Hence, there is no discussion here of the ways in which knowledge outside a domain influences learning. It is a fruitful topic, but outside the scope of this thesis. The domain or subject matter knowledge possessed by a learner may be of several types.

Declarative knowledge refers to content information the learner possesses. While it encompasses specific facts, it is more than that as it also involves the full range of conceptual schemata a person has for a particular concept. The richness of those schemata varies considerably from person to person, a point which I will consider shortly. Declarative knowledge in any area can be general domain knowledge (Alexander et al., 1994). Domain knowledge is comprised of one's general knowledge of physics, chemistry, or the geosciences, to use science examples. A person could also possess topic knowledge (Alexander, 2003). One's knowledge of kinematics in physics or glaciers in geoscience would be examples of topic knowledge. Topic knowledge might be even more specific such as types of glaciers or the velocity differences that occur at various points within a glacier. Even experts in a field do not possess detailed knowledge of all subtopics within their field. In fact, people are usually experts on one narrow topic within a domain. Nonetheless, those experts generally possess a rich framework of domain knowledge within which their expertise resides.

It is possible to have quite extensive topic knowledge in a narrow area but have little domain knowledge. This would be unlikely to describe an expert in a domain, but it could describe someone who is a novice in the field. Consider the example of an amateur fossil hunter to illustrate how a novice might possess topic knowledge but more limited domain knowledge. The collector may be familiar with several good fossil-yielding sites and recognise a large number of specimens by sight. Yet, that same individual may have little knowledge of what a particular fossil indicates about the depositional environment or the paleoecology of the region.

As its name implies, procedural knowledge means knowing *how* to do something. Like declarative knowledge, procedural knowledge exists at the domain and topic levels. Domain procedural knowledge has to do with the ways of “doing” a particular discipline. History is not done in the same way as physics or chemistry. There are different procedures for the investigation of phenomena in each of those disciplines. One only becomes a member of the guild by learning the way the discipline is “done.” There are also more specific or topical procedures. The method for taking strike-and-dip is a geoscience example that is often learned in an undergraduate field geology course.

The distinctions between what constitutes declarative and what is procedural knowledge are not always clear. Principles of stratigraphic correlation are examples of geoscience procedural knowledge in that they provide a set of guiding criteria that are used to correlate strata in the field. On the other hand, before those ideas are used procedurally they are generally learned as a set of axioms that introductory geoscience students must recite on an exam. Doing well on the exam does not guarantee the student will have an easy time using them in the field. Block diagrams in geology textbooks are a long way from actual outcrops. Nonetheless, the declarative knowledge becomes a stepping stone to the procedural. On the other hand, as one gains procedural experience in the field, that declarative knowledge becomes more solidified and more meaningfully known within a richer set of schemata. In that sense the declarative and procedural inform each other.

Self-regulatory knowledge deals with how a learner understands himself and his knowledge of the domain in question. It includes how the student monitors his own thinking and problem-solving within a domain (metacognition). Self-regulatory knowledge also relates to attitudes toward a particular discipline and how a learner views himself as a student of that discipline. If a student sees himself as incapable of mastering a particular domain (say maths or chemistry), there will be implications for future learning within the domain.

Just as declarative and procedural knowledge inform one another, as the stratigraphic correlation example illustrates, each component of subject matter knowledge interacts with and influences all the others. General domain knowledge provides a framework for the acquisition of specific topic knowledge. Topic knowledge, like that possessed by the amateur fossil collector, could provide a stepping stone to more robust domain knowledge. Self-regulatory knowledge affects how a student approaches new information whether it be declarative or procedural. At the same time, the extent and nature of one's declarative and procedural knowledge impact self-regulatory knowledge.

Within this scheme, we must now ask what role prior subject matter knowledge plays when new information is encountered. One neoclassical constructivist view (Chi, 2008) proposes that existing ideas are part of students' mental models of how the world works. In this view prior knowledge may serve as a building block for new knowledge, but it can also hinder new learning, as what may seem to be perceptually or intuitively correct turns out to not be so. New scientific information is sometimes added to current conceptions even when the new information is actually unrelated to existing knowledge. From Chi's perspective, this occurs if the individual has no prior knowledge of the topic although the person may possess some related domain knowledge. When a student possesses some correct but incomplete prior knowledge, new information serves to "fill in the gaps" and is added to existing schema as was true when the pupil had little or no prior knowledge. It is equally possible that the student's current "knowledge" contains one or more alternative conceptions that have arisen from everyday experience or from formal, school learning. In this scenario, the new information conflicts with what the learner already *knows*. Therefore, learning requires conceptual change. This is a useful framework within which to probe how current knowledge affects concept acquisition more specifically.

Section 1.3 briefly outlined the lack of a consensus definition in the literature of the term *concept.* The lack of a common definition is a major obstacle in any attempt to synthesise literature on subject matter knowledge and its role in concept acquisition (K. Murphy & Alexander, 2008; White & Gunstone, 2008). If we are fuzzy on just what a concept is, trying to explain the role of current knowledge in the acquisition of new concepts is a formidable task. Many science concepts are of the latter type and fit nicely within this framework. A rich conceptual understanding of a domain in science contains declarative and procedural knowledge that is both broad and deep. The self-regulatory knowledge an individual possesses helps her use the knowledge she already has and acquire more knowledge in the domain. We would expect an expert in a field to have broad and deep understanding, although even an expert's understanding of some topics will be deeper than others.

2.4.2 How experts and novices use subject matter knowledge

One line of research has contrasted how experts and novices (and sometimes those in-between) approach new learning in a particular domain. There are several varieties of this type of research. In some studies two groups, experts and novices, are studied (e.g., Chi, Hutchinson, & Robin, 1989). Others have expanded the experimental design to include more groups to better reflect a view of learning as a continuum (e.g., K. Anderson & Leinhardt, 2002). My focus in this section is primarily on the novice end of the continuum. The reality is that very few students we encounter in compulsory education or those in introductory courses in U.S. universities are on their way to becoming experts in a given field (Alexander, 2003). I will distinguish characteristics of experts from novices, but this discussion is designed to set the stage for how students who are not experts in a domain reason when they lack significant subject matter knowledge.

Novices are novices in a field precisely because they possess less declarative, procedural, and self-regulatory knowledge than experts in that field. The domain or topic knowledge they do possess tends to be isolated rather than part of a rich schema of ideas (Petcovic & Libarkin, 2007). One expert-novice study related to declarative knowledge involved 10 five-and six-year-old children who were experts or novices about dinosaurs (Chi et al., 1989). Children were placed in groups based upon their scores on a pretest of their dinosaur knowledge. The authors use a category membership definition of the term *concept* applied to dinosaurs similar to Murphy's (2002) mentioned earlier, although that does not diminish its applicability to the present study in any way. In one task, children were shown pictures of unfamiliar dinosaurs (those that do not generally appear in children's books about dinosaurs) and asked questions about where they might live, what they would eat, how they would defend themselves, and what other dinosaurs they might be related to. When making inferences about unfamiliar dinosaurs, novices based their inferences on general knowledge of animals ("like a rhinoceros," p. 45). Those inferences tended to focus on surface features (i.e., a dinosaur was like a rhinoceros because they both have a horn). In contrast, experts based their inferences upon knowledge of families of dinosaurs as well as individuals. They were much more likely to say something like, "It's a meat-eater because it has sharp teeth," (p. 50). They were aware of which specific features were important for category membership. Their relatively richer conceptual schemata provided them with a frame of reference that permitted them to attend to the features that were truly important to the task at hand. Novices, who only possessed a frame of reference related to animals in general, had no real way to judge which features were critical to answer the question.

Novices also have less procedural knowledge. Anderson and Leinhardt (2002) explored the ways in which individuals on the expert-novice continuum approached problem-solving geography tasks involving maps. There were 30 participants in their study: seven expert geographers, seven described as novice geographers (enrolled in their first cartography course), seven referred to as advanced novice geographers (geography majors, who had completed at least two cartography classes), and nine preservice (trainee) secondary social studies teachers who were midway through a master's program. The latter were preparing to teach social studies to students in grades 7-12 (ages 12-18).

All participants were individually interviewed using a map. They were asked to draw a line to indicate the shortest distance between two cities: New York and Moscow, New York and Cape Town, Santiago and Singapore, Anchorage and New York, and Anchorage and Santiago. A world map was used for the first three questions and a map of the Americas was used for the last two. In order to complete this task correctly, a person needed to know that sizes and distances are distorted at the poles on a Mercator projection map. This is the type of map that is most commonly found in textbooks or in classrooms as a wall map. Due to this distortion at the poles, the shortest distance between any of the two cities would not be a straight line but a curved one. Furthermore, the depth of the curve is determined by the distance of the cities from the poles.

Several findings are interesting. Even the experts did not draw all lines correctly. There are several possible explanations for this. First, this could be due to the way answers were scored. To be scored as correct, the line drawn by the participant had to be within one inch on either side of the correct line for the distance between the two cities. While inter-rater reliability for scoring was high (95%), that is expected since the criterion is stringent and fairly straightforward. The high inter-rater reliability could give the appearance that the scoring criterion was a valid measure of a person's competence at the task. Yet, the size of the map or the scale could have played a role in a person's accuracy. Both of the maps used can be described as small-scale maps in which map distances equate to large differences on the ground. What is not clear from the authors' description is the size of the map. A wall-size world map has a greater margin of error than a textbook-size one would. If a person was off by one inch on the wall-size map this would represent a smaller distance on the ground than that one inch error would represent on a textbook-size map showing the same geographic area. Another possibility is that this is similar to what Chi (2006) describes as experts' failure to attend to details or display overconfidence. It does not represent any error in understanding but may reflect the idea that this line is "about right." If those same experts had been asked to draw a line that would indicate a flight route for an airplane they may have drawn the line more carefully. This is speculative, of course, but it is a plausible inference.

Second, as would be expected, participants with greater domain knowledge drew more accurate lines than those with less knowledge. Individuals preparing to teach secondary social studies scored the worst. This is unfortunate since these individuals will be teaching basic map skills to students one day. However, there is nothing in the study to indicate that these students had ever taken a cartography course or been exposed to the type of task they were asked to complete. In that sense, this group was more "novice" than those who were designated as novices by Anderson and Leinhardt. Many of the future teachers did not appear to realise that a two-dimensional map distorts distances from a globe as they made reference to the notion that "the shortest distance between two points is a straight line" (p. 308). However, some of those preservice (trainee) teachers made comments indicating they realised there would be distortion from the globe to the map but they did not know how to draw a line to reflect that fact.

Third, novices employed a variety of general reasoning strategies to determine the solution to the problem. In the same way that children who knew little about dinosaurs (Chi et al., 1989) used general knowledge of animals to answer questions about novel dinosaurs, participants in this map study employed general problem-solving strategies when they lacked domain-specific ones. Sometimes these reasoning strategies were quite robust. One novice used a visualisation strategy in which he imagined what he terms a "half-way map" on which lines of longitude converged as they do on a globe (p. 304). This strategy helped him solve three of the five tasks successfully. He did, however, possess some domain knowledge since he knew that lines of longitude are not parallel, and he knew that piece of information was important for solving the problem. The authors do not discuss which of the tasks this student solved successfully and which ones he did not. That information would be helpful in shedding further light on the extent to which he was able to apply his domain knowledge.

All groups in this study applied pre-existing knowledge to the task. Those with greater domain and topic knowledge were better able to determine *what* knowledge was important. When they lacked relevant domain or topic knowledge, people resorted to more general knowledge and reasoning strategies. Sometimes that led them astray as in, “The shortest distance between two points is a straight line.” At other times, their strategies were powerful enough to enable them to achieve a measure of success on the task.

Alexander and Judy (1988) conducted a literature review of studies dealing with the relationship between domain knowledge (either declarative or procedural) and strategic (self-regulatory) knowledge. They use the term *strategic knowledge* to denote an idea that is similar to what I have referred to as self-regulatory knowledge. They define strategic knowledge as a type of procedural knowledge, which makes their sense of the term somewhat different from how I have used self-regulatory knowledge. However, the authors refer to strategies as “goal-directed” and aids for “regulation, execution, or evaluation” of a task (p. 376). Thus, although their meaning for the term *strategic knowledge* is not identical to the way in which I have defined self-regulatory knowledge, there is enough similarity in the terms to apply Alexander and Judy’s comments about strategies to this section on self-regulation. I will use the term self-regulatory knowledge in the interest of consistency, with full recognition that the term itself does not appear in their paper but the idea does.

They report a number of findings from their literature review. As was seen for declarative and procedural knowledge, there is interaction between self-regulatory knowledge and the other types. General self-regulatory knowledge that transcends domains, as well as that which is specific to a particular domain, helps students both use the declarative and procedural knowledge they currently possess and acquire more knowledge. In turn, self-regulatory knowledge changes as students learn more about a domain. Students who do not possess a lot of domain knowledge, but do have well-developed general self-regulatory knowledge apply that general logical reasoning to novel tasks and are often successful. Further, novices who successfully solve problems in a domain seem able to see the underlying similarities between two problems that appear at the surface to be quite different from one another. I will return to these last two points shortly.

Novices or experts in a domain and those in-between are regularly encountering new information in that area. Their present subject matter knowledge also influences how they evaluate this new information. Chinn and Brewer (1993) state that if students are presented with anomalous data that contradicts their existing ideas, students with little background knowledge are more easily convinced by the anomalous data than those with more knowledge. If the new information is correct and the present knowledge it conflicts with is minimal, having less rather than more background knowledge is advantageous. On the other hand, students with less subject matter knowledge possess little basis on which to critically evaluate data. They may not even recognise that this new data conflicts with what they already know. They may judge that information is credible when it is not.

This rather extended discussion of expert-novice literature relates to the discussion in section 2.4.1 but also transcends it. Thus far, I have considered how knowledge within a specific subject area such as the geosciences or chemistry influences learning. However, students do not only bring knowledge of the subject matter in question to the learning task. They also bring broader knowledge of other domains and still more general knowledge and ideas about how the world works that cross domains. This is probably most obvious for self-regulatory knowledge which involves strategies for approaching novel tasks as well as a student's general metacognitive abilities. It includes dispositions and attitudes toward learning in general and how someone views herself as a learner. As was discussed earlier, novices who have well-developed self-regulatory knowledge are sometimes able to compensate for their lack of knowledge by their general problem-solving skills.

Similarly, declarative and procedural knowledge exist outside a particular domain. If a learner has little knowledge in one area she will only be able to relate new information or experiences to knowledge she has in some other domain. The dinosaur study described earlier (Chi et al., 1989) demonstrates this point. Children who knew little about dinosaurs drew analogies between an unfamiliar dinosaur and an animal with which they were familiar (rhinoceros). To use another example, suppose I know little about dogs or procedures about how to care for dogs, but I do have experience with horses. If I am given a dog, I will have minimal "dog knowledge" on which to relate to my new pet. I may not know what dogs eat, how to groom them, what kind of medical care they need, or that they need to be walked regularly. I will have some "horse knowledge" that I will draw upon to help me with my new dog. There are some similarities between dogs and horses since both are mammals. Yet, my dog is likely to be quite unhappy if I feed him hay and oats for dinner. What if I have no "horse knowledge" but instead I have "goldfish knowledge?" Now my dog is in a much more difficult position as there are fewer commonalities between dogs and goldfish than there are between dogs and horses. My "goldfish knowledge" does not help me very much with taking care of my dog.

The point here is that background knowledge provides some sort of framework or referent by which new information is judged and categorised. From a constructivist perspective there is a relational quality to knowledge. Without a sense of how pieces of information fit together, knowledge is isolated and inert. Second, if a learner possesses minimal or no background knowledge in an area, problem-solving relies on the knowledge he does possess so that there is a greater reliance on common surface features rather than deep structure.

The ability to create meaningful referents may be a key factor that separates novices from experts in a field. Jones and Taylor (2009) describe interviews with 50 professionals from a wide variety of fields from scientists (of many types) to engineers to artists to a chef. All deal with scale (size and/or distance) regularly as part of their profession. This study relates to those described in section 2.3.2 about the role of a unit, but it nicely illustrates an important characteristic of experts in a domain. Thus, it is discussed here rather than in the earlier section. Participants were asked to reflect back upon in-school and out-of-school experiences they felt were beneficial in developing their understanding of scale. Repeatedly, these individuals described the use of anchor points [I have been using the term referents] against which to judge the size of other objects or distances between objects. As was mentioned in section 2.3.2, there is a relationship between numerical units and specific subject referents that epitomise those units. The subject knowledge informs the numerical knowledge and vice versa.

2.4.3 Summary of research on subject matter knowledge

The subject matter knowledge a learner brings to a task is multi-faceted. It is composed of declarative—content knowledge, procedural—"how-to" knowledge, and self-regulatory—reflective knowledge. All affect whether and how new concepts are learned. Declarative and procedural knowledge can exist at the more general domain level and also at the more specific topic level. Self-regulatory knowledge occurs at the domain level.

The amount of knowledge a person possesses plays a role in what happens when new information is encountered. Because novices in a domain lack coherent, structured knowledge within that area, they must rely on knowledge in other areas. The fact that they don't know much about the important ideas in a particular domain means that they often rely on surface features to determine what "related" knowledge to use. Sometimes those decisions mean that new information is connected to existing ideas to which they have little to no conceptual relationship. On the other hand, people who successfully solve problems in a domain in which they are novices often have good self-regulatory knowledge that cuts across domains and makes up for what they lack in domain or topic knowledge.

2.5 A framework for the review of deep time conceptions research

Before reviewing the research on students' conceptions regarding deep time, it is essential to ask how the literature reviewed in previous sections might be germane to the topic. This provides a framework for the review that will follow.

2.5.1 Conventional time concepts and a concept of deep time

How are succession and duration understood as part of a conception of deep time? This is a major part of the empirical investigation of this thesis. Are these two ideas as important to an understanding of deep time as they are to conventional time? Are they understood in a deep time context in ways that are different from how they are understood in the context of conventional time? I will first discuss succession as an issue in a conception of deep time and then duration.

2.5.1.1 Succession in deep time

Succession involves the before and after relationship and is essential for an understanding of deep time. One is asked to look at an outcrop and infer the process(es) that resulted in what one now sees. Principles of stratigraphy such as cross-cutting relationships (igneous intrusion or fault is younger than strata it cuts across), original horizontality (sediments are deposited in basically horizontal layers), and superposition (in an undisturbed sequence of strata the oldest layer is at the bottom) are employed to determine the succession of events that led to a particular formation and require one to work backwards to sequence those events. Ault (1982) and Dodick and Orion (2003a, 2003b) allege that Piagetian type research into succession in conventional time differs from deep time in its reliance on physics motion problems while geology requires one to interpret static sequences. It is certainly true that Piaget saw a close connection between motion and spatial distance and time. Much of his methodology that was described in section 2.2.1 employed phenomena from physics, to be sure. Nonetheless, his conclusions about the underlying notions of what constitutes a concept of time are very germane to geologic time since those static rock sequences required motion of some type to become what they now are. Several of his tasks cannot be criticized on the basis of the observance of physical motion (see section 2.2.1).

Furthermore, children do not merely use motion and spatial distance to determine duration, but also use starting and ending times (Levin, 1982; Levin, Iraeli, & Darom, 1978). The problem is that in the case of geologic strata neither starting/ending times nor the visual perception of motion are available to help someone determine the sequence of the strata. In some instances this is not at all a straightforward process and there is considerable disagreement amongst experts regarding the course of events that produced a particular formation. This does not imply that a concept of deep time is somehow qualitatively different from conventional time and succession in deep time works differently from succession in conventional time. It merely indicates that an understanding of succession alone is insufficient for the task. The fact that perceptual information available in the conventional time tasks is not available when judging succession in deep time does not in and of itself make the two fundamentally different. If it can be shown that judging succession is equally difficult in the absence of similar perceptual information for conventional time tasks, this would suggest that something other than a qualitative difference between conventional and deep time is at work.

Another complicating factor is the need to conceive of the succession of events on an immense time scale. In relative terms, this involves placing the extinction of the dinosaurs prior to the appearance of humans on that scale. An understanding of absolute succession requires the knowledge that the amount of time from the extinction of the dinosaurs to the appearance of humans is considerably less than the amount of time from the formation of the Earth to the appearance of the dinosaurs.

2.5.1.2 Duration in deep time

A concept of duration as it relates to deep time is bound tightly to a notion of the rates of geologic processes. The concept of duration as inversely proportional to velocity is applicable in the same way it is critical to a notion of conventional time. A student must be able to dissociate size from rate and realize that size alone cannot be equated with duration. Two layers of sedimentary strata may be the same thickness but have been deposited at very different rates, and, hence, represent different durations. Similarly, a layer of shale may underlie a layer of thicker volcanic ash. In this case, the thicker ash layer would have been deposited in a shorter time than the thinner layer of shale. Many of the studies discussed in section 2.2.3 employed tasks that were highly analogous to what is required to interpret a stratigraphic sequence. For example, an object was placed some distance from a starting point and the participant had to infer the amount of time necessary for the object to traverse the distance (Acredolo et al., 1984; Berndt & Wood, 1974; Matsuda, 2001; Wilkening, 1981). The key difference between these tasks and the geologic ones is that participants in the conventional time studies were likely to know something about the rate at which a turtle and a cat move. They may not be so likely to know much about the rates of deposition for sedimentary strata. That hypothesis is supported by the fact that in one study (Casasanto & Boroditsky, 2008) university undergraduates did have difficulty inferring durations (see section 2.2.3). Adults in the other studies in which they were part of the sample had virtually no problems with the task. This could be explained by the fact that the one task was less familiar to them than the others.

In thinking about rates of geologic processes one must not only think about very long periods or units of time, but one must also conceive of processes that happen at very slow rates and produce significant aggregate effects. The growth rate of the Himalayan Mountains is an example. Currently the Himalayas are growing at a rate of about 1 cm per year. At this rate, they will be ten kilometres taller than they are now in one million years. If a child were to visit the Himalayan Mountains today and then return in 60 years, no size difference would be perceived. Perception itself will not help a person comprehend a growth rate of 1 cm/year since within a human lifetime there will be no observable difference to that person viewing them at a 60-year interval.

An understanding of duration also requires a clock in which units of time progress independent of the events they measure. The thickness of a layer of rock does not necessarily indicate the length of deposition. When measuring time by a clock, size and speed are immaterial. The only thing that matters is the passage of those independent units of time.

2.5.1.3 Summary of succession and duration in deep time

In my view, deep time is *quantitatively* different from conventional time, but it is not *qualitatively* different. It requires the same basic concepts; however, they must be applied to a context that is outside human ability to directly experience. Those two concepts are succession (one event precedes another in time) and duration (the amount of time required for an event to occur is inversely proportional to the rate at which it occurs). The literature on how consistently adults apply ideas of duration and succession to conventional time is a bit murky. The reasons for the inconsistencies are even more puzzling. If a student does not comprehend conventional time he or she is unlikely to understand deep time. To argue that the two are qualitatively different requires that there be a different mechanism to account for them. If what separates them is the fact that one occurs within a human lifetime and one does not, then we must ask what about a concept of 1,000 years or even 500 years? Both of these also occur outside a human lifetime and yet are not viewed as qualitatively different from time periods that are within a human lifetime. If the difference is qualitative at what point does it become so? Is it at the onset of recorded history, or perhaps the appearance of *Homo sapiens*? The difference is one of degree. This leads us to ask how conceptions of large numbers and geoscience content knowledge may influence an understanding of deep time.

2.5.2 Large numbers and a concept of deep time

It may seem obvious at this point that a poor conception of large numbers will impact a student's understanding of deep time. We have seen that numbers appear to be mapped spatially in a manner that exaggerates differences closer to the reference point and minimizes those farther away. If students are unable to discriminate between two large numbers, lack a meaningful referent against which to judge their size, and have little concept of large units of time, deep time is likely to be a great mystery. For example, if I have neither a referent nor a conception of a one million year unit, it will simply be a very large number that may not appear much different from a one thousand year unit. A good conception of proportional relationships amongst numbers in the base-ten system contributes to an understanding of the difference between a one thousand and a one million year unit. Consider how differences across orders of magnitude change the scale. One thousand seconds ago was 17 minutes ago; one million seconds ago was 12 days ago; and one billion seconds ago was 12,000 days or 33 years ago.

This issue is further complicated for time because the entire scale is not built upon a base of ten. Arabic numbers and linear sizes are. In order to move from one to another from very small to very large requires one to multiply and divide by powers of ten. When dealing with time across the entire scale, one encounters base ten at the extremes (fractions of a second, 100 years), but units of time encountered on a daily or yearly basis are not in base ten at all. Thus, the proportional relationships are even more complicated for time than they are for linear distances.

2.5.3 Geoscience content knowledge and a concept of deep time

The amount and nature of geoscience content knowledge a student brings to the learning task also influences how deep time is understood. A certain amount of declarative knowledge is required to make sense of geoscience processes. In order to have a sense of the amount of time necessary for collisional mountain ranges to form, it is useful to know something about plate tectonics and the average rate at which plates move. If someone thinks the Appalachian Mountains were formed in a period of hundreds of years, it could be because the person is unaware that tectonic plates move at the rate of centimetres per year or the individual is unable to use the information to infer anything about the time required for a collisional mountain range to form. In this case declarative geoscience knowledge interacts with an understanding of units of time. Even if the student knows something about the rate of tectonic plate movement, centimetres/year may have no meaning as a unit.

Domain and topic knowledge in the geosciences impacts how people understand deep time in other ways. Suppose a student is asked to place the Appalachian and the Himalayan Mountains on a relative time scale to indicate which is older. Some topic knowledge about the two mountain ranges could be useful. More general domain knowledge of how weathering and erosion change mountains over time would also help a student complete the task. If the individual lacks relevant topic or domain knowledge, the person must rely solely on more general knowledge. This could lead a student to apply faulty reasoning to the task. While not universally true, there are many instances in the natural world where size and age are positively correlated. The biological world is a good example as adults of a species are generally larger than juveniles. If a student lacks knowledge of how surface processes change mountain ranges, that student may incorrectly conclude that the Himalayas are older than the Appalachians simply because they are larger. A lack of geoscience content knowledge means a student focuses on a surface feature (size) which leads to an erroneous conclusion.

I mentioned principles of stratigraphic correlation in section 2.5.1.1. There I made the connection between those principles and the way succession is judged in conventional time. They are also related to a student’s declarative and procedural subject matter knowledge in the geosciences. If a student does not use fossil succession to sequence a series of strata containing fossils, it does not necessarily mean the learner has a poor understanding of temporal succession in deep time. Rather, the student may simply be unaware of fossil succession and not know how to use index fossils to correlate strata in two outcrops.

2.6 Literature on conceptions of deep time

A review of the literature suggests that students of all ages as well as many in-service teachers hold similar alternative conceptions regarding deep time (Orion & Ault, 2007). It is well established that at least some understanding of deep time is essential to conceptual understanding of various geoscience processes (Hume, 1979; Zen, 2001).

There are three basic lines of research. The first focuses specifically on the ability of subjects to place geoscience events in correct temporal order and to indicate specific dates for the events. The second finds questions about deep time located within more general geoscience conceptions research. The third explores how students understand processes that occur in deep time. It will come as no surprise that the first two are concerned with succession rather than duration. However, the third type is also primarily concerned with succession. In fact, there has been minimal attention to conceptions of duration in deep time.

Rather than organise this section by those three types, I have chosen to group them by whether they deal with succession or duration. Research results from all three types of studies that relate to succession will be discussed first, followed by those dealing with duration. The ways in which results can be understood in the light of conventional time concepts, large numbers, and geoscience knowledge will be described.

2.6.1 Research on succession in deep time

One of the earliest researchers to look specifically at how students understand the temporal order of geologic processes was Charles Ault (1980; 1982). One part of his investigation involved a task in which 2nd, 4th, and 6th grade children were shown three plastic tubes containing "core samples" from various locations in a hypothetical compost pile in which each successive layer spread out completely over the previous layer (consistent with Nicholas Steno's principle of lateral continuity). The children's task was to determine the sequence of layers in the compost pile and explain their reasoning. Subsequently, the children were shown three diagrams of rock exposures from the local area, locations which Ault says were familiar to most of the children in the study. The children were then asked where the oldest rocks could be found in each of the outcrops. Even those who were able to satisfactorily sequence the layers in the compost pile were less successful determining relative age of the layers in the geologic formations. Children who correctly deduced that the oldest layer was at the bottom often did so for the wrong reason. In his 1982 paper, Ault posits that the primary school children in his study were unable to apply temporal reasoning to geological settings because the geological phenomena themselves were unfamiliar to the students. This inference is exactly what would be predicted in light of what was discussed in section 2.4. Ault reports the reasons children cited for why the oldest rocks in the formation would be found in a particular location. The children's rationales indicate they were relying on surface features in the absence of specific geoscience content knowledge. Some used colour, while others thought "crumbliness" was the most salient feature to determine age. Although there is nothing in Ault's writing to indicate this, it is quite possible that children have noted the difference in concrete sidewalks or walls that are newer versus older. Older ones are often characterised by broken or "crumbly" pieces of varying sizes. While this is clearly speculative as there is no evidence to support my claim, it would not be surprising if children applied similar reasoning to the rock sequences in the field. Further, some children in his study said the oldest rocks were found in the centre. This accretionary view also fits with certain everyday experiences such as making a paper maché mask. The innermost layer of paper and paste was applied first and is, therefore, the oldest layer of the mask. Ault concludes,

> *Children, and by extension, adults, who are ignorant of geological concepts cannot grasp the meaning of geological time, not necessarily because the number of years involved is large, but because their time conceptualization has no referent in the rock record,* (1982, p. 309).

Ault further argues that his geologic task is fundamentally different from Piaget's with its reliance on static sequences versus Piaget's use of motion. However, Nersessian (2008) points to research that shows people can perform mental spatial transformations such as rotation and reflection on both two-and three-dimensional figures. In other words, they can take a static image and move it mentally. Thus the static versus motion distinction may not be so important. This would seem to contradict Ault's assertion that methodological differences between his research and Piaget's are at least partially responsible for why the children in his study were unable to determine where the oldest rock layers would be found. More likely, the children simply didn't know there was any reason to imagine the strata being deposited one on top of the other according to the principle of superposition. They made no connection between the outcrops and the compost pile activity because they had no reason to think the two were connected. Ault's contention that the children's lack of geoscience content knowledge played a role in the children's responses is a better explanation.

A second aspect of an understanding of succession in deep time is how individuals perceive temporal order at that scale. Roger Trend (1998, 2000, 2001a, 2001b) provides an interesting line of research in this arena. Trend used similar methodologies which he varied slightly in his work to determine deep time conceptions held by 10-and 11-year-olds, primary teacher trainees, and primary teachers in the U.K. An additional study with 17-year-olds was a bit different from the others and is discussed separately. He used several different tasks, but only the ones that are germane to this thesis will be discussed here. His first study was conducted with 177 10-and 11-year-old children. Each child individually sorted three sets of cards containing either eight or ten events from geologic history. There was some overlap of items on the three sets of cards so they could be used to establish reliability. Children were only required to place events in relative order. However, the pilot study on which this investigation was based did yield information about children's ideas about an absolute age for the Earth.

In the studies involving primary teacher trainees (2000), conceptions of deep time were assessed in several different ways. Eighty-five primary trainee teachers completed a questionnaire containing 20 geo-events from the Big Bang to the first appearance of humans on Earth. Participants were provided with nine temporal categories and asked to indicate which time period corresponded to when each event occurred. Time periods ranged from “less than one thousand years ago” to “more than a million million years ago.” An additional 60 primary trainee teachers completed one of two card sort tasks in which some respondents were given a set of 21 cards containing the same 20 geo-events as those in the questionnaire and a final card labelled “present day.” In the first task, participants merely sequenced cards in temporal order. In the second task, respondents were given an identical set of cards (different colour) but they were also provided a sheet showing the geologic time scale divided into 40 divisions with “present day” as the most recent. The final two divisions were named “one million million years ago” and “older than this: please state,” (p. 545). In a follow-up study, in-service primary teachers (2001b) completed the same questionnaire the 85 primary teacher trainees did.

Primary teacher trainees performed more accurately when placing events in relative order as opposed to assigning absolute ages to them. The trainee teachers sorted events into three broad categories: extremely ancient, less ancient, and geologically recent (Trend, 2000, p. 552) in contrast to the 10-and 11-year old children in his first study who only sorted into two categories: extremely ancient and less ancient (1998). Thus, the trainee teachers appeared to have a more well-defined concept of long time periods than the children did.

Eighty-two percent of the primary teachers placed the formation of the Sun prior to the Big Bang based upon their absolute time rankings. Amongst trainee teachers, even though rank order of events for the groups with and without the geologic time scale were similar, those receiving the geologic time scale tended to use the entire scale and assigned dates greater than 50,000 million years ago to events from the Big Bang through when the first rocks were made on Earth. These answers are well outside the bounds of current scientific understanding which places the Big Bang at around 13.7 billion years ago. In the later study some in-service primary teachers placed even more events prior to the onset of space-time (Trend, 2001b) while others placed those early events as less than 1,000 years ago. Although the main study with 10-and 11-year-olds only dealt with relative time, the six children in the pilot study who answered the question, "How old is the Earth?" gave answers that ranged from the thousands to one trillion years old.

What could explain the use of dates that are off by many orders of magnitude? Trend says, "The huge numbers involved appear to cause confusion," (Trend, 2001b, p. 212). This makes perfect sense in light of the research reviewed in section 2.3.1. If an individual does not possess a good conception of the distinctions between thousands and millions or millions and billions, it will be difficult to deal with the numbers necessary to understand deep time. In his first paper on the topic, Trend notes a similar issue regarding the role an understanding of large numbers could have on conceptions of deep time for younger students.

> *Not surprisingly for 10- and 11-year-old children, their conception of large numbers becomes a dominant influence on their answers. They indicate little more than wild guesses, designed to convey the idea of an immense period of time,* (Trend, 1998, p. 980)

Trend's study with 17-year-olds (2001a) employed a different methodology than the other three studies. First, students were asked to list as many events from Earth's history as possible. They were instructed to try to include a variety of events representing life, climate, surface features, landforms, and Earth materials. An additional group of students was given a set of 24 cards, 18 of which contained the name of a geo-event and six of which contained a date. Students were asked to construct a concept map with the cards and write statements that linked the cards to each other. Thirty-six students who completed the first task provided unsolicited information about temporal or causal links between events. Nine of those attributed Earth's formation to the Big Bang. Concept maps also indicated a rather poor understanding of the temporal or causal relationships between events. Specific events were linked with dates that were off by many orders of magnitude. Further, students appeared to meld a few key events in their minds and link them in a causal fashion. For example, 24 participants linked the Ice Age with dinosaurs, most often as the cause of their extinction.

Trend posits that learners bring a mental representation of a set of key geo-events that serve as a framework for how they understand relative and absolute succession in deep time. This deep time framework (DTF) acts as an Ausubelian advance organizer for new learning in geoscience. Trend says,

> *A "deep time framework" (DTF) is proposed as the learner's personal chronology of key geo-events, including their relative and absolute dates, which they bring to bear when encountering new geo-events or geoscience phenomena. It is used by learners to assimilate new learning and the quality*

> *and security of that new learning is influenced by each learner's existing cognitive framework* (Trend, 2001b, p. 192).

These key geo-events are those which serve as benchmarks around which to place other geo-events. He distinguishes between an individual's own DTF and a curricular DTF which would consist of decisive geo-events that would increase future learning and lists possible items for inclusion in a curricular DTF.

An important component of a deep time conceptual framework is the idea of the succession of events on an immense time scale. In relative terms, this involves placing the extinction of the dinosaurs prior to the appearance of humans on that scale. An understanding of absolute succession requires the knowledge that the amount of time from the extinction of the dinosaurs to the appearance of humans is considerably less than the amount of time from the formation of the Earth to the appearance of the dinosaurs. The notion of a framework implies that people possess an overarching model for deep time that is at least somewhat coherent.

Several themes appear in his writing that relate to research described in earlier sections of this chapter. The first is the role that an understanding of large numbers plays in people's ability to estimate ages in deep time. Compression of the timing of events is common. Confusion between the timing of the Big Bang and the formation of the Sun or the formation of the Earth has already been mentioned. It might be argued that the introduction of a time scale that included dates prior to the Big Bang in several of the studies predisposed participants to the notion that those dates must be appropriate for some of the geo-events; however other research suggests that similar student responses are obtained in the absence of such a scale. In one study

(Libarkin et al., 2007) where no such time scale was provided, university students' responses to, "How old is the Earth?" ranged from thousands of years to trillions. In another study (Janssen et al., 2006), the range was thousands to hundred billions. In Marques and Thompson (1997) the range was hundreds of years to trillions.

A second theme is Trend's contention that within a deep time framework certain key geo-events serve as benchmarks around which to place other events. This is reminiscent of literature on large numbers that was described in section 2.3.1 which showed that when placing numbers on a number line, children rely on reference points and then place numbers in relation to those referents.

Trend (1998, 2000, 2001b) is well aware of the role geoscience content knowledge plays in responses. In his initial study with 10-and-11-year-olds, he reports high standard deviations for the relative placement of the Big Bang and the Ice Age in the card sort. He explains this deviation for the Big Bang by suggesting that some children have never heard of the event and, hence, place it randomly (Trend, 1998, p. 985). In a later study (2000), primary teacher trainees also exhibited considerable disagreement about the placement of the Ice Age. Once again, Trend posits that this event holds no temporal or evolutionary meaning for respondents. Therefore, its placement in a sequence is speculative.

In his 2001 study with primary teachers, Trend describes his rationale for including the item, "trilobites became extinct." He expected a high standard deviation on this item since it was likely to be less well known than other items in the questionnaire, and, indeed, that was the case. Interestingly, its relative position was fairly close to where it should have been. Trend suggests that respondents may have been using cues from other events to figure out where to place this unknown event, a highly plausible inference. The placement of the earliest geo-events in this study had the lowest standard deviations, although there was some confusion as to whether the Big Bang preceded or followed the formation of the Sun. In contrast, the later geo-events had the highest standard deviations. This could be explained by the notion that in the absence of the age of certain events such as the formation of planet Earth, it is possible to reason that it must have occurred early in the sequence since many of the other events require an existent Earth for them to take place. However, it is not so easy to reason where “Atlantic Ocean started to open,” or “Woolly mammoths became extinct,” belong in the sequence without some knowledge of Earth’s history.

Another study employed a technique very similar to one of the tasks in the interview protocol for this thesis. Libarkin, Kurdziel, and Anderson (2007) had 63 university students from four U.S. universities place four events on a timeline that began with “Earth forms” and ended with “Today”: appearance of first life on Earth, appearance of dinosaurs, disappearance of dinosaurs, and appearance of humans. All students were enrolled in an introductory biology or geology course at the time data was collected. (As a reminder to the reader, introductory science courses in the U.S. include students from a variety of majors not simply those who are studying science.) Thus, students come to these courses with varied exposure to and interest in science topics. Students were judged on the relative placement of events. A few (six) students indicated that dinosaurs and humans co-existed, but overwhelmingly students placed events in the correct order. However, very few students plotted events in relative positions that would agree with scientific consensus on the placement of events in time. Specifically many placed the appearance of dinosaurs much closer to the appearance of first life than would be appropriate. This could well indicate a lack of knowledge of evolutionary biology and the timescale it requires. It could also mean that the numbers on that timescale have little meaning to students.

Trend's work demonstrates that students are quite confused about absolute dates for geologic events. Other research substantiates his findings. Catley and Novick (2009) asked university students from two U.S. institutions to provide dates for seven key events in evolutionary history: origin of the Earth, first fossils, eukaryotic cells, Cambrian explosion, first mammals, dinosaur extinction, and the appearance of *Homo habilis*. These questions were part of a larger questionnaire that provided data for several studies. The 126 participants in the study were deemed to be of stronger or weaker background depending upon the amount of previous biology or geology coursework they had taken. Stronger background students were on average one semester farther along in their education than the weaker science background ones and had taken more science courses. On average the stronger background students reported 3+ semesters of biology or geology courses. This is contrasted with an average of 0.5 courses for the weaker background students. Estimates for the age of events ranged across seven or more orders of magnitude for each event with no clear differences between students of stronger versus weaker background with one exception. Stronger background students were more likely to underestimate the dates of events, which the authors term "forward telescoping." A similar phenomenon was observed by Janssen, Chessa, and Murre (2006) and described in section 2.2.2 in which people tended to underestimate the age for events that occurred more than 1,000 days ago. Unlike the stronger background students, the weaker background students in the Catley and Novick study assigned dates that were more evenly distributed across a range of times. Compression of the timing of events was common amongst these students although some compressed events farther back in time while others compressed them closer to the present.

Of the 126 participants in the study, a total of 16 (7 stronger and 9 weaker background) did not provide age estimates for any of the seven events. Due to the nature of the experimental design, there is no evidence that these students were asked why they did not attempt the task. There are many possible reasons from disinterest to a perceived inability to do so. The number of courses taken could not have predicted who the non-responders were likely to be.

Additional research that deals with both relative and absolute succession sits within the context of studies that probe students' conceptions across a range of geoscience topics. They primarily collected data via written responses, although several paired written responses with at least some interviews (Dahl, S. Anderson, & Libarkin, 2005; Libarkin et al., 2005; Marques & Thompson, 1997; Rule, 2005). Many of the written responses were forced choice formats, either multiple choice or true/false. However, there were several which used some or all open response items (Dahl et al., 2005; DeLaughter et al., 1998; Libarkin et al., 2005).

When 10-11-year-old and 14-15-year-old pupils were asked about the origin of the Earth, Marques and Thompson (1997) found that a number of participants equated the formation of the universe with the origin of the Earth, findings that are consistent with Trend. About half of the children also indicated that the appearance of life on Earth occurred at Earth's formation. This could be evidence for a compression effect—old events are all simply old events. The differences between them are not discriminated. There is some hint of an increasing ability to deal with larger numbers with age. Fourteen-and-15-year-old children were much more likely to indicate an older age for the Earth than 10-and-11-year-olds.

The issue of the numbers themselves appears in at least one of these more general geoscience studies. Oversby (1996) compared geoscience conceptions of primary and secondary school children with pre-service postgraduate student teachers. Participants were asked a series of questions, some of which required a yes-no response and others of which were open-ended. In response to the question, "How old do you think the Earth is?" Oversby says,

> *A minority held the accepted scientific view that the Earth is older than 10^9 years but the results could be partly explained by assuming that the respondents had a poor understanding of large numbers and an inability to distinguish between millions and billions. Occasional answers which simply said 'millions and millions' were testament to this interpretation,* (p. 95).

Taken together, these general geoscience studies indicate confusion regarding both absolute and relative succession. There is generally not enough information provided to speculate what is beneath those incorrect ideas. Some general findings are:

1. Dinosaurs lived contemporaneously with cavemen (Schoon, 1992; Schoon & Boone, 1998).

2. Life existed before (DeLaughter et al., 1998; Marques & Thompson, 1997) or at Earth's formation (Libarkin et al., 2005).
3. A supercontinent was present at Earth's formation (Libarkin & S. Anderson, 2005).
4. The Earth is anywhere from hundreds to millions of years old (Marques & Thompson, 1997; Oversby, 1996).
5. Carbon-14 is the preferred method to date the age of the Earth (Dahl et al., 2005).
6. Coal formed at the same time as the Earth (Rule, 2005).
7. Students' understanding of geologic time after instruction is not appreciably different from their understanding prior to instruction (Libarkin & S. Anderson, 2005; Libarkin et al., 2005).

Dodick and Orion (2003a, 2003b) have been interested in how students understand geologic processes in time, both in terms of succession or temporal order and duration or lengths of geologic processes. I will provide a brief description of their methods here and then discuss items in their instruments that relate to succession. Those that relate to duration will be described in the next section (2.6.2). They distinguish between what they term a passive temporal framework which involves a sequence of events in absolute time and an active logical understanding which involves reasoning about the relative time of events, the relationship of adjacent strata to one another, and how geologic changes occur over time. In their view, all the deep time research reviewed thus far in this section is of the first type. Dodick and Orion's attention to active, logical understanding is important. A solid understanding of deep time requires the ability to use relative dating principles to order events or geologic strata, as well as the ability to situate events in their appropriate places on a timeline. They base their work on the ideas of Jacques Montangero (1996; Pons & Montangero, 1999) who was discussed in section 2.2.4. I have already attempted to show why I do not believe the notion of diachronic thought is a useful construct to think about either conventional or deep time. However, my disagreement with them on this point in no way diminishes the value of their research. Dodick and Orion's work provides very interesting findings that are important when considering how students understand deep time. Furthermore, the manner in which I have organised this discussion demonstrates that their findings can be interpreted in the light of succession and duration.

The authors developed a questionnaire, the Geological Time Assessment Test (GeoTAT) which consists of a total of seven line drawings and associated questions related to relative dating principles. An additional drawing showed a room in which a crime had obviously recently been committed. This item was included to see whether students would approach a sequencing problem differently when geoscience content knowledge was not a factor than when it was. A number of the puzzles required students to place events or rock layers containing fossils in temporal order. Sometimes those rock layers had been folded so that a fossil could be below another one in the drawing but represent a more recent depositional event. In other cases, there were unconformities or gaps in the rock record. A layer containing a particular fossil might be present in one column of strata but missing from another. Students had to use information about fossils and/or rock types to infer depositional environments in several other puzzles. In another puzzle, students used an outcrop with rock layers containing leg and foot fossils demonstrating horse evolution from *Hyraotherium* to a modern horse to deduce the events that led to the outcrop depicted in the drawing. All of the puzzles in the GeoTAT involved relative age except one. That puzzle had an outcrop with two igneous layers alternating between three sedimentary layers containing fossils (Figure 2.3, p. 132). Students had to provide ages for the sedimentary layers. This puzzle will be described more fully in the next section.

The GeoTAT was specifically designed to limit the geoscience content knowledge required to complete the tasks, thereby only assessing students' temporal understanding. They did this in two ways: first, by providing some information explicitly, and second, by limiting the relevant content information to things students would likely be familiar with. More will be said about this point shortly.

As part of its development the GeoTAT was tested with 156 Israeli eleventh and twelfth graders (2003b), some of whom were studying geology and some of whom were not. In Israel it is possible for high school students to major in a subject, which explains why some but not all of their sample were geology students. Some of the non-geology students were studying biology and would likely have had some exposure to evolutionary biology. Geology students generally outperformed non-geology students, although differences were not always statistically significant. There was also an age effect, as 12th grade geology students performed better than 11th graders, a fact which the authors attribute to the cumulative effect of fieldwork. An age trend was not so consistent with non-geology students. In some instances, 12th

graders outperformed 11th graders, but in other instances they did not (Dodick & Orion, 2003b).

In the main study (2003a) the GeoTAT was used with 285 seventh through twelfth grade Israeli students. None of these students were studying geology at the time of the study. Not surprisingly, ninth through twelfth graders were significantly better able to interpret geologic processes and phenomena than seventh and eighth graders. However, older students did not always score better than younger pupils, i.e., 12th graders didn't always do better than 11th graders. The authors do not speculate about why this may be the case. The fact that seventh and eighth graders were less successful at solving the puzzles is consistent with what would be expected based upon research on conventional time described in section 2.2.2 (e.g., Friedman, 2005). The authors hypothesize that seventh or eighth grade is the lower limit at which a student can display actualistic thinking, i.e., the notion that in geoscience the present is the key to the past.

Despite the fact that the authors attempted to minimize the effect of geoscience content knowledge on responses, they are nonetheless aware that it is extremely difficult to do so. The fact that 12th grade geology students generally outperformed 11th grade geology and 11th and 12th grade non-geology students in the validation study indicates that geoscience content knowledge must be a factor in responses, a point which the authors clearly acknowledge (2003b). Results from the main study further illustrate how geoscience content knowledge may have affected responses.

One task (Puzzle 5) consisted of three outcrops each containing five layers with the names and pictures of fossils found in those layers. They showed students the picture in Figure 2.2 with the following instructions, "*The illustration below represents three rocks exposures containing fossils. Try to order the fossils according to their implied age, from the oldest fossil to the youngest fossil*" (Dodick & Orion, 2003a, p. 439)

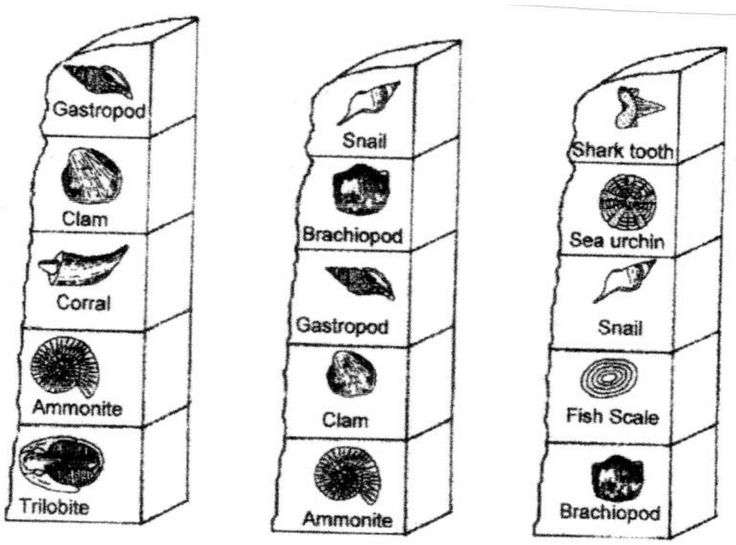


Figure 2.2 Puzzle 5 of the GeoTAT (from Dodick and Orion, 2003a, p. 439)

This could be expected to be an extremely difficult task for anyone who knows little about fossil succession and how it is used to correlate sedimentary strata. To answer the item correctly, a student would need to know that two layers of rock that are separated from one another but contain the same fossil are the same age. Additionally, they would need to understand that sometimes there are unconformities (gaps) in the rock record due either to erosion or an absence of deposition. One unconformity exists in the second stratigraphic column in Puzzle 5 between the layer containing the ammonite and the one containing the clam. The missing layer is present in the first column (the coral). Dodick and Orion felt that students who were able to use principles of temporal organization correctly would be able to figure out how to sequence the strata correctly. However, specific geoscience content knowledge seems to be a prerequisite for the successful completion of this task. Without knowledge of relative dating principles (declarative and procedural knowledge), a student would have to base an answer on other factors. The correct order of deposition is: trilobite, ammonite, coral, clam, gastropod, brachiopod, fish scale, snail, sea urchin, and shark tooth.

The authors note that it was common for students, particularly the 7th and 8th graders, to order the fossils randomly. They say, "Indeed, it was not rare to see such students order three different fossils from three different localities temporally equivalent because they appeared on the same stratigraphic level (2003a, p. 428). The authors may be correct that students' orders were random. Indeed, they *appear* random to someone versed in the geosciences. However, there may be an alternate possibility. If I am to complete the task and know little or nothing about fossil succession, I must reason on some other basis if I wish to answer the question. I will need to rely on any bit of general knowledge about the world that I possess and can retrieve at that moment. Concluding that two fossils that occupy similar positions in a column are temporally equivalent is not an unreasonable inference if I have little else upon which to base my answer. Students were not asked why they ordered fossils in the way they did so my next statement is speculative. It is entirely possible that students were reasoning that in a pile, the thing on the bottom was put there first and things on top were added later. Thus I may not be ordering the fossils randomly at all, even though I am using an incorrect strategy (in this instance). Of course, if an outcrop contains a series of undisturbed strata in which there are no unconformities, then reasoning that the oldest layer is on the bottom is correct. For Puzzle 5, that reasoning is not sufficient. I must know *something* about fossil succession, even if I don't know the term, in order to complete this task successfully.

Grade 9-12 students performed significantly better on this task than 7th-8th graders, although the highest mean score for any grade was 55%. The authors hypothesise that this could be due to an improvement in visual-spatial ability. A follow-up experiment in which they presented four of the GeoTAT puzzles and 14 visual-spatial tasks to 172 10th-11th graders found a strong correlation between the two sets of tasks, thereby lending support to their hypothesis. Given the relationship between space and time described in section 2.2, it is not at all surprising to see such a correlation. Even though older students scored better than younger pupils, a mean score of 55% is not high. Although visual-spatial abilities may account for the improvement with age, an absence of knowledge of principles of stratigraphic correlation may explain why no groups scored particularly well on the task. Without the relevant geoscience content knowledge, a student is left only with whatever knowledge or experience she has to reason her way to an answer.

A second example of the pivotal role played by geoscience content knowledge is found in Puzzle 6c. The question says,

> *When scientists excavated this area deeply they found an alternating arrangement of layers consisting of marine sedimentary rock containing no fossils and terrestrial sedimentary rock containing fossils of dinosaurs. What is the significance of this alternating arrangement of layers containing terrestrial*

> *sedimentary rock containing dinosaurs, and marine sedimentary rock without dinosaurs?* (Dodick & Orion, 2003a, p. 728)

To answer this question correctly a student first needs to understand the meaning of the terms marine and terrestrial. That understanding is not a given. Secondly, the student would need to know something about a transgressive-regressive sequence. Such a sequence is evidenced by vertical strata that indicate rising and falling water levels in a shallow sea. Sediment types that are normally deposited adjacent to one another are deposited atop each other leaving a record of rising and falling sea levels. The presence of layers containing dinosaur fossils interspersed with marine sedimentary layers indicates just such a sequence. However, this would not likely be clear to a student who has not studied historical geology or evolutionary biology. The authors state that this sequence indicates the migration of dinosaurs in and out of the region in response to changing sea levels. The authors interpret difficulty with this puzzle as indicative of problems with interstage linkage (borrowed from Montangero). That is to say, students don't understand cause and effect relationships or that some events are prerequisites for other events. That certainly seems to be the case; however, I would argue that it is not due to any failure to comprehend time. Rather problems with this puzzle are indicative of a lack of geoscience content knowledge. It seems to me that the extent of a person's knowledge of transgressive-regressive sequences is the best explanation for their findings. That claim is further substantiated by the fact that students who participated in a program called "From Dinosaurs to Darwin," a curricular unit integrating evolution and palaeontology all improved dramatically on this question from their pre-unit to post-unit performance (Dodick & Orion, 2003b).

Hidalgo and Otero (2004) looked at both succession and duration. Their duration task will be discussed in section 2.6.2. Two tasks were used to probe understanding of succession. They administered a written test to a group of 16-year-old secondary school students who were enrolled in an elective natural sciences course and 19-20-year olds who were in a post-secondary school technical programme. Students were asked to list events of "personal, historic, anthropologic, paleontologic, geologic, or cosmologic character" (p. 849) that occurred in eight time periods ranging from 10 to 100,000,000 years ago. Individuals of both ages were less able to place events in the correct temporal category the farther from the present the events were located. This result is, of course, consistent with the research on both conventional time and large numbers that was reported in sections 2.2 and 2.3. The farther from the reference point, the present in this case, the less likely a student was to be able to place it in its correct temporal category. Participants were subsequently shown four pictures each showing a different moment in evolutionary history.

- Invertebrate marine life against a hilly background containing no life forms
- Two dinosaurs in a flat landscape with some vegetation
- A variety of mammals and birds on a varied landscape
- A group of hominids in a forested area

Pupils were asked to place the pictures in the correct temporal order and assign a temporal label to each picture. It does not appear that they were provided with a range of dates as was true in the first task. Rather, it seems students had to come up with their own temporal labels. While over 80% of the older students placed the pictures in the correct order, none applied the correct temporal label to a picture. In other words, they could order the events on a relative scale but had no idea how long ago each occurred. Fifty-six percent of the younger students ordered the pictures correctly and a few did correctly assign temporal labels to the pictures. The authors propose that some students in their study may not have actually known the temporal order of the pictures. Rather, they were able to deduce the correct order by *using other knowledge* [emphasis mine]. For example, one student noted that the picture containing marine life had no terrestrial vegetation in the picture. Therefore, the student concluded that terrestrial vegetation couldn't survive in that environment (or it would have been in the picture). Hence, the picture with terrestrial life must come after the picture with marine life. This student appears to have possessed sufficient knowledge of evolutionary biology to know that life appeared first in marine environments and only subsequently on land. The authors conclude,

> *...the high percentage of incoherence among the arrangements of events and labelling of antiquity suggests that the arrangement strategy used by the students on that level depends more on deduction than on the recovery of memorized labels* (Hidalgo & Otero, 2004, p. 854)

2.6.2 Research on duration in deep time

All of the studies just reported, including those dealing with geoscience processes, have involved succession. Dodick and Orion (Dodick & Orion, 2003a, 2003b) also investigated duration. Puzzle 4 of the GeoTAT, which was described in the previous section, is interesting in terms of what it indicates about conceptions of duration. This was used with the same group of 285 seventh through twelfth grade pupils described earlier. Students were shown the drawing in Figure 2.3.

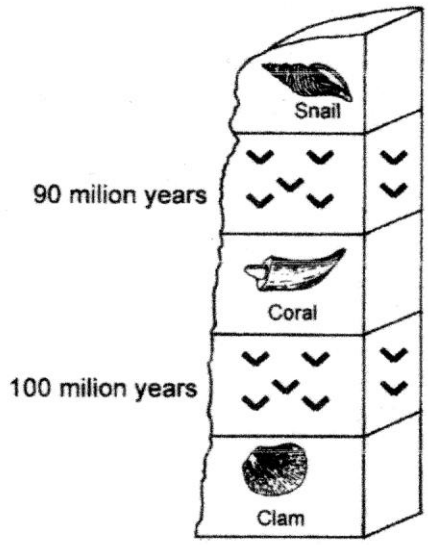


Figure 2.3 Puzzle 4 of the GeoTAT (from Dodick and Orion, 2003a, p. 439)

They were told that layers of igneous rock (denoted by V) lay between the sedimentary layers containing fossils (clam, coral, and snail). The igneous rock layers had been dated by geologists in the lab and their ages were written on the drawing. Students were asked to determine the absolute age in years of the fossil layers. The correct answer for the clam would be >100 million years. The coral would be between 90 and 100 million years, and the snail would be <90 million years. Ninety-eight percent of the students in the sample (at all ages) were unable to correctly solve this puzzle. The most common error was to judge that each rock layer represented the same amount of time, thus the snail layer was judged to be 80 million years old and the clam layer was 100 million years. This is very reminiscent of some of the younger children in Piaget's tasks who equated amount of work done with duration. In fact, these students seemed to be working from the premise that equal size means equal durations. Dodick and Orion have provided good evidence that the equating of spatial size and time operates in deep time much as it does in conventional time. A person's geoscience content knowledge comes into play as well. If I know little about sedimentary or igneous processes, I am at a loss when trying to determine what information from this drawing is important to date the sedimentary fossiliferous layers. It is not unreasonable for me to base my answer on the size of the layers. Since the strata appear to be the same thickness, the conclusion that they all required the same amount of time to form seems to be a logical inference.

In addition to the GeoTAT, Dodick and Orion designed an additional test called the Strategic Factors Test (SFT) that sheds additional light on how students understand duration in deep time (2003a). They administered this test to a group of 52 Grade11-12 students who were all majoring in earth science. Therefore, they might be expected to possess at least some of the requisite geoscience content knowledge necessary to complete the task. Participants were shown three drawings each containing a pair of rock outcrops with strata of different thicknesses and number of layers. They were asked to compare two layers and determine which was older, if possible. They were told there were four possible answers:

- A is older
- B is older
- both are the same age, or
- can't be determined (Dodick & Orion, 2003a, p. 422).

In one situation both outcrops had the same number of layers but Outcrop B was taller. In that scenario, the layers in B were thicker than the ones in A. In the second situation both outcrops were the same height, but A had twice has many layers as B. In the final scenario, A had twice as many layers as B and B was taller than A. In all cases, the correct answer is that it is impossible to know which layer is older without knowing something about depositional rates (the salient piece of geoscience content knowledge).

Table 2.1 lists the percentage of students who correctly answered that it was impossible to determine from the drawing alone which of two layers was older. Only a minority of students answered each question correctly. Perhaps even more telling is the fact that less than 10% of the students who answered correctly said anything about rate of deposition in their explanation. Some answers could be considered partially correct, as there were students who said that information regarding rock types or environmental conditions was missing. While environmental conditions are a factor in rates of deposition, this answer was deemed incomplete since how environmental factors affect rates of deposition was not mentioned.

Equal number of layers, B is taller	Both layers same height, A has twice as many layers as B	A has twice as many layers as B, B is taller
35%	31%	29%

Table 2.1 Percentage of students who correctly answered *impossible to know* which layer is older on SFT (Dodick and Orion, 2003a, p. 432-433)

Students who did not answer correctly often thought that the more compact the rock layer the older it was. Compaction was often cited by people in the second situation who said a layer in Outcrop A was older. Yet, when both varied the number of layers was a more salient factor for students than the size of the outcrop. In situation three, about 75% of those who were incorrect said that the layer in Outcrop A was older.

On the surface, it may appear that students are employing two different strategies to judge durations in these examples, however, they may actually represent similar ideas. It can be argued that size is equated with duration in both instances. In the first one because the total number of layers is the same, the size of the entire outcrop determines which is older. Therefore, taller = longer duration. On the other hand, when the heights of the entire outcrops are the same, size is not spatial distance but the number of layers from the starting point. The reasoning is that it takes longer for more layers to accumulate. Therefore, the number of layers indicates age. When both vary, number of layers is viewed as more important than total height of the outcrop. Interview data indicates that students relied upon what they knew about sedimentation as a process of compaction to make this judgment. They were making some use of their geoscience content knowledge. Since their knowledge was incomplete, they reached an erroneous conclusion. Thus, errors may not be indicative of problems with time per se but rather insufficient domain or topic knowledge.

Hidalgo and Otero (2004) investigated concepts of duration using a series of written questions. In one, students were shown a picture of a fold caused by flexion and asked to indicate what may have caused the fold. The majority of students attributed the fold to a catastrophic process such as an earthquake, a volcano, or a plate collision. The authors state that students who indicated tectonic plate collision could be the causal agent did not appear to view the process as requiring a long interval of time. Unfortunately, they do not provide information about how many students attributed the fold to lithospheric processes but over incorrect time scales since all catastrophic changes appear together in their data. Further, they do not include information about what time scales those students did assign to them. About the only thing that can be said is that their results appear to mirror those of others who found that students have a poor understanding of the amount of time necessary for geologic processes to occur (Libarkin et al., 2007).

2.6.3 Summary of the literature on conceptions of deep time

Several observations can be made from the literature on conceptions of deep time. The most obvious one is that succession has been explored far more frequently than duration. This is not at all surprising since it is especially difficult to design ways to investigate duration that are not highly sensitive to geoscience content knowledge. Much of my remaining comments thus deal with succession. Students of all ages and practicing teachers hold very similar concepts regarding deep time. Age and prior coursework in biology or geology appear to have only a minimal effect on participant responses. Age of respondents seems to be correlated with an older view of the Earth but sometimes this means that people state that the Earth is much older than it actually is.

There are hints that poor discrimination amongst orders of magnitude exists. There is some evidence that individuals compress events farther back in time compared to more recent ones. Equating the Big Bang with the formation of the Earth or the extinction of the dinosaurs are examples. Libarkin et al., (2007) state that university students in their study placed more time between the appearance and disappearance of terrestrial dinosaurs than is warranted while compressing the time between the appearance of the first life on Earth and the appearance of the dinosaurs. This could be explained by positing that people don't know and simply offer their best guess on the matter. On the other hand, these findings are very similar to Confrey's (1991) interview with Suzanne described in section 2.3.1. There may be little discrimination between such large numbers. A million and a billion are both simply really big numbers. The difference between the two is unclear. Hidalgo and Otero (2004) state, "the subjects of our study frequently used very short intervals of time in geologic terms," (p. 854). They also indicate that participants used labels such as one million years but seemed unaware of the amount of time represented by those labels. The use of a compressed logarithmic scale mirrors what we have seen for both large numbers and also for events in conventional time, e.g., Janssen and Chessa (2006), and is consistent with the spatial example of the two hills in the distance described in 2.2.5.

There were only two studies reviewed that dealt with duration at all. Yet, the latter is just as important to an understanding of deep time as the former. In fact, it may be impossible to truly have a sense of the immensity of geologic time without some sense of the rates and durations of geologic processes. Dodick and Orion's work suggests that adolescents equate size with duration in a geologic context, much as the younger children in Piaget's work equated duration with either distance travelled or amount of work done. It is difficult to say much from Hidalgo and Otero other than to repeat that students frequently underestimated the time required for geologic processes to occur. It is important that students' concepts of duration be explored more fully than they have to date.

Overall understanding of deep time is poor, but we are left to speculate why that is the case. Most researchers report responses as correct/incorrect, but we have only hints as to the type of thinking that separates the two groups. Further, we do not know much at all about the students who give correct responses. In cases where students were asked the reasons for their responses, it is clear that correct responses may be coincidental more than anything else. Therefore, we must explore *why* deep time is so difficult to grasp. Only one researcher (Ault, 1980, 1982) attempted to probe children's understanding of both conventional and deep time. No one has conducted a similar study with older students. Additionally, the literature is replete with suggestions that geoscience content knowledge or an understanding of large numbers may be influencing how students respond to questions about deep time. However, the extent to which those two variables might explain students' difficulties with the concept has not been explored.

2.7 A model for a concept of deep time

Anyone who has taught geoscience at any level of the educational system has seen how difficult it is for students to comprehend geologic time. The fact that many students make very modest gains in their understanding after instruction (Libarkin & S. Anderson, 2005) suggests that we may need a clearer picture of what is beneath the surface of their misunderstanding if we are to have any hope of addressing this issue within an instructional context. If we are to move beyond simply describing what is, then it is useful to begin to ask why it is so. What factors underlie a conception of deep time? In this chapter, I have attempted to show that the difficulties students have with deep time mirror the difficulties they have in other areas. Based upon that research base, I propose a working model to account for the key underlying factors that are necessary to comprehend deep time. Research reviewed in sections 2.6 alluded to what those underlying factors might be. This model arises out of the literature that has already been reviewed in this chapter as well as some preliminary interviews conducted prior to the main study for this thesis (see section 3.3). I contend that the concept of deep time can best be represented by the metaphor of a "three-legged stool" (Cheek, in press). The three "legs" are the pieces that support this understanding: conventional time, large numbers, and the subject matter knowledge of the geosciences (geoscience content knowledge). The result, or the "seat" of the "stool," is a solid concept of deep time. As a stool needs all three legs to be sturdy and bear weight, so an understanding of deep time requires all three of these "legs" to be in place. A three-legged stool might have legs of equal size, but it is also possible for its legs to be of unequal length and yet be stable. At present no evidence exists to suggest whether the three "legs" of the "deep time stool" bear equal weight or whether any of these factors is more critical to a conception of deep time. Thus, it could be that a person with well-developed concepts of conventional time and large numbers, but only limited geoscience content knowledge is able to grasp deep time to some extent. The way(s) in which the "legs" interact is equally unclear. Does growth in one area foster growth in the others or do they develop independently? Despite the unanswered questions, the model enables us to explore student understanding in the three areas and then see how that knowledge is applied to a geologic context. A poor understanding of deep time may reflect:

1. Failure to comprehend conventional time or failure to apply notions of conventional time to deep time. Students do not have a solid grasp of succession or duration. They are unable to judge them accurately in everyday situations. If they are able to deal with succession and duration in conventional time, they act as if deep time operates by a different set of rules than conventional time does.

2. Poor understanding of large numbers. Deep time differs from conventional time by many orders of magnitude and employs quantities rarely encountered in the course of daily life. Students possess poor number sense for numbers of great magnitude and minimize the differences between them. They lack reference points and have trouble conceiving of temporal units that are outside their everyday experience.

3. Limited geoscience content knowledge. Alternative conceptions regarding deep time are based on ideas that have little to do with time itself. Students' minimal knowledge of geologic processes means that they are reasoning on the basis of everyday ideas that may not coincide with scientific ones.

2.8 Other factors that influence the stability of the "stool"

I have posited a "three-legged stool" to account for a conception of deep time. Yet, there are other factors that influence a learner's ability to comprehend deep time. An exploration of these factors is outside the bounds of this thesis.

Nonetheless, they must be acknowledged as factors that may affect how at least some students understand deep time.

The first is the role of an individual's metaphysical beliefs upon a conception of deep time. References to a student's belief system as it influences their understanding of deep time appear in the literature, particularly in the U.S. (e.g., Catley & Novick, 2009; Libarkin et al., 2007). In fact, the issue of people rejecting deep time on the basis of deeply held beliefs appears to be more of an issue in the U.S. than in other countries as it is not mentioned as a factor in research conducted outside the U.S. In some cases, individuals may have a solid understanding of conventional time, a good sense of large numbers, and the requisite geoscience content knowledge to understand deep time as a concept yet reject the notion of deep time as being a fundamental characteristic of the world. This rejection has less to do with conceptions and more to do with whether they feel acknowledging that deep time exists is in conflict with their fundamental views about the world, often their religious beliefs. Chinn and Brewer describe how people tend to ignore, reject, or reinterpret anomalous data when it conflicts with well-entrenched beliefs (Chinn & Brewer, 1993). One could argue that just because a student rejects the notion of deep time doesn't automatically signal a lack of understanding. Conversely, it might be said that if a person truly understood the concept it wouldn't be rejected. In the case of strongly held beliefs, I would argue that it *is* possible to understand deep time, but reject it nonetheless. In a sense the distinction may not matter. Students, who reject the possibility of deep time on the basis of metaphysical beliefs, do not use the concept to reason appropriately about the timing or duration of events in Earth's history. In essence, the concept doesn't exist for them.

The credence a student gives to media representations of the past also affects a learner's conceptions. Many television programs and films portray humans and dinosaurs as cotemporaneous. If a student accepts these sources as accurate portrayals of the past, a concept of the sequence of events in deep time will be distorted.

A third factor relates to a student's attitudes and motivation. This relates to their role in the acquisition of a concept, which is clearly important. In university systems like the U.S. where students are sometimes in a course simply to fulfil a requirement rather than for any interest in the topic, it can clearly be a factor. Second, students' attitudes and motivation play a role in how they respond to the questions in our instruments. The inferences we make as researchers about student conceptions rest upon the assumption that participants are accurately expressing the conceptions they have which may, in fact, not be the case. It is a point we must be aware of when drawing conclusions from our data.

2.9 Summary and conclusions

We are once again brought back to the model of the "three-legged stool." I have attempted to show how each of the three "legs:" conventional time, large numbers and geoscience content knowledge contributes to a conception of deep time. I have also attempted to link each of those "legs" with published research results. These ideas are not new. Many people working to uncover student conceptions of deep time have acknowledged the role played by each of these "legs." What has not yet been attempted is a systematic attempt to investigate these "legs" *together in a single study with the same group of learners*. This thesis seeks to explore each of these three "legs" in turn to determine how they might affect a conception of deep time. The first research question explores the relationship between how learners understand conventional and deep time. This investigation probes understanding of succession and duration at both conventional and deep time scales. This permits a comparison between how students understand those ideas at different time scales. If students have a poor understanding of conventional time, they would perform poorly on succession and duration tasks in conventional time even if the tasks themselves "looked" geological. This study also represents an attempt to help redress the imbalance between succession and duration studies in the literature.

The second research question concerns students' understanding of large numbers. If larger numbers are a significant factor in a conception of deep time, we would expect that students would have difficulty with tasks dealing with numbers alone apart from a geologic context. Their performance on "pure number" tasks would be no better than their performance on tasks that involve geologic events.

Finally, the third research question asks whether learners' understandings of deep time are at least partially due to the amount of content knowledge they possess. If students' difficulties with deep time can be partially accounted for by a lack of geoscience content knowledge, we would expect to see several things. If we are able to remove content knowledge as a significant variable, we would expect that students would experience little difficulty with tasks in which that knowledge is usually a factor.

Second, if students are unfamiliar with events and processes that occur in conventional time, they should be no better at judging succession or duration than for events in deep time.

This exploratory study is designed to test the working model. We now turn to the specific research design.

CHAPTER THREE

METHODOLOGY OF THE RESEARCH

The purpose of this study is to investigate whether the deep time “stool” with its three “legs” is a viable working model to account for students’ difficulties understanding deep time. First, I discuss the rationale for the methodology I chose. Strengths and weaknesses of the methodology are outlined. Next, I describe the preliminary interviews that led to the development of the instrument, and then, the items in the instrument itself. Finally, I provide information about the sample.

3.1 A pragmatist approach to quantitative and qualitative research methods

The debate about the relative merits of quantitative versus qualitative research has raged for decades (Lancy, 1993). Decisions researchers make about methodologies reflect their worldviews or paradigms (Mertens, 2005). An exposition of the various paradigms that underlie educational research with their accompanying strengths and weaknesses can be found elsewhere (Creswell, 2009; Lancy, 1993; Mertens, 2005).

There is great value in quantitative research with its emphasis on theory testing and isolation of variables to determine the distinct role they play in students’ conceptions. One of the strengths of quantitative research is that its samples are often relatively large. Samples are selected in such a way as to represent the population about which one wishes to generalise. The appropriate quantification of data allows for statistical manipulation of that data and the ability to compare results from one study to those of another. Quantitative research is often favoured by government agencies in the U.S. (Denzin & Lincoln, 2005; Silverman, 2000) because it is seen as being valid, reliable, and less influenced by the researcher's own perspective than qualitative methods. Theoretically, quantitative data analysis is less dependent upon the researcher's personal views on the subject (Libarkin & Kurdziel, 2002). Validity and reliability are based on statistical analyses and, thus, objectively determined. However, no research study is value-neutral but always reflects the underlying philosophy and values of the researcher.

On the other hand, Silverman (2000) points out that quantitative research is sometimes characterised as narrowly focused and thus divorced from a real world context. The need to operationalise definitions can sometimes result in the very arbitrariness that quantitative researchers seek to avoid. Choices must be made when constructing definitions, and it is difficult to completely remove researchers' assumptions and values from the process. Ambiguity must be resolved. This can result in the collapsing of data into one category when if distinct categories had been created and data from each category had been analysed separately, the researcher might have reached different conclusions. Decisions about which variables warrant study and how data are analysed will always reflect an investigator's overall orientation (Lancy, 1993). Reliance on quantification means that only certain types of questions are asked—those that can be represented numerically and subjected to appropriate statistical manipulation and analysis. People who engage in quantitative research are not ignorant of its imperfections. High quality quantitative research designs acknowledge the limitations of the various quantitative methodologies and do all that is possible to mitigate them, sometimes by the inclusion of qualitative methods at some point in the process.

Qualitative research is often portrayed as a more holistic examination of students' ideas in more naturalistic settings (e.g., Marshall & Rossman, 2006). However, this distinction may serve to imply qualitative research is "good" while quantitative is "bad" as the latter often occurs in "unnatural" settings. Those who engage in qualitative research represent a variety of underlying paradigms from those who see the researcher as independent from the individuals studied, to others who see themselves as participants with the subjects of the research, to those who believe the role of their research is to advance a particular political or social agenda (Creswell, 2009; Mertens, 2005). Because it encompasses many different traditions and underlying paradigms, qualitative research can be more difficult to characterise. Lancy (1993) compares qualitative research to a "mixed forest" (p. 3) in which the various traditions sit almost independently of one another. Even though similar methodologies are employed, the assumptions upon which they rest differ substantially.

A primary strength of qualitative research is the richness and volume of data it generates. Some might view the richness of the data as a drawback since data analysis can be quite time consuming. Qualitative researchers would argue that what can be gleaned from the data is well worth the time required for its analysis. These investigators would also say that the fact that data analysis is so closely tied to the data source is a second strength of the methodology. In some cases this means that study participants can validate or modify conclusions (Silverman, 2000). Data collection in qualitative studies often evolves as the study progresses making these methods flexible to changing circumstances or to new information that is uncovered early in the study. New questions are asked, while others are jettisoned to generate data with greater depth and breadth. This stands in contrast to quantitative methods in which data collection methods are standardised. Sometimes qualitative research is viewed as inductive while quantitative is deductive. That may be another example of applying terms that are unfair to both traditions, since it appears to set up a false dichotomy. Qualitative researchers employ both inductive and deductive means of data analysis (Hoepfl, 1997), and quantitative researchers do as well.

There are some clear drawbacks to qualitative methods. The relatively small sample sizes make it virtually impossible to make defensible generalisations about populations. Further, samples are not chosen in such a way as to be representative of the larger population. Instead, qualitative samples are often purposefully chosen to elicit the views of particular individuals or groups (Hoepfl, 1997). Thus, while the sample chosen may reflect the range of responses one would find in the population as a whole, it is unlikely that those responses would occur in the same proportions in the population as they do in the sample. All three of these concerns make generalising very problematic. Qualitative researchers sometimes use the term *transferability* rather than *generalisability* (Marshall & Rossman, 2006) to describe the relevance of a study's findings to another context. On occasion, this can translate into a lack of concern for the need to make one's own findings transferable to other situations. Marshall and Rossman state, "The burden of demonstrating that a set of findings applies to another context rests more with the researcher who would make that transfer than with the original researcher" (2006, p. 201-202). Thus, in their view, a researcher is not responsible to show how her study's findings transfer to a larger population. That task falls to those who read her results and want to apply them to a new situation. The researcher is only responsible to provide enough detail so that the reader can make an informed decision about whether the findings are transferrable or not. This leads to one of the strongest criticisms levelled at qualitative research—its anecdotal nature (Silverman, 2000).

Definitions of reliability and validity are different in qualitative and quantitative research (Creswell, 2009; Hoepfl, 1997; Libarkin & Kurdziel, 2002). To those who are more comfortable with quantitative definitions, qualitative measures of reliability and validity can appear "fuzzy." In fact, sometimes those terms are not used at all. Instead, validity is referred to as *credibility* while reliability becomes *dependability* (Mertens, 2005).

One way to increase reliability or dependability has to do with the consistency of coding of responses. Having more than one person code data and reporting inter-coder agreement can provide readers with important information to judge the study's dependability. Additionally, research studies need to provide clear information about how responses were coded. When readers are given information about how coding was done and examples of responses that were coded in particular ways, they can better evaluate how consistently coding decisions were made. Quantitative research seeks to not only establish reliability in terms of how responses are judged but also the consistency of the responses themselves. There is not the same assumption in qualitative research that responses will necessarily be consistent over time (Mertens, 2005). Thus, qualitative research is more concerned with the first type of reliability but far less with the second.

One way to increase a study's credibility or validity is to triangulate data or collect data on the same idea in multiple ways (Marshall & Rossman, 2006). If responses are indicative of actual conceptions or themes we would expect to see evidence of that across tasks. If they are not, this discrepancy casts doubt on their validity or credibility as indicators of subjects' ideas. An additional point is the need to assess a wider variety of related ideas in the same study. Johnson and Gott (1996) say, "A criticism we have of much of the research into children's thinking is that studies have tended to be compartmentalized and have not sought to develop an understanding of a child's responses over a range of related ideas" (p. 567). That is precisely the point of this study: to assess students' thinking across related ideas to determine how they connect to one another.

The criticism that qualitative data is often anecdotal has already been mentioned. It is easy to choose a few "good" examples of student responses that illustrate categories yet not attempt to deal with data that is less clear or contradictory (Silverman, 2000). Thus, when writing up qualitative research it is important to include data that is confusing or that seems to be anomalous (e.g., Creswell, 2009; Mertens, 2005). Qualitative data reports in which all responses appear to fit neatly into categories have probably left out important pieces of information. The inclusion of detailed descriptions permits a reader to draw his own conclusions about the credibility of the investigator's conclusions.

My own research orientation falls somewhere between the two extremes on a qualitative-quantitative continuum. This study is best described as one that fits into a pragmatist paradigm (Creswell, 2009). Within this view, decisions about research methodologies are not based upon a commitment to a particular method as the "right" way to conduct research as this limits what can be learned. The strengths and weaknesses of each methodology must be evaluated in light of the research question in order to determine the appropriate methods to best answer that question.

Pragmatism is frequently associated with mixed methods research that utilises both quantitative and qualitative methodologies (Mertens, 2005), although pragmatists would not say that *every* study should use both types of methodologies or that they should be applied in equal measure across studies. If we accept that quantitative and qualitative methods each have a role to play, as pragmatists do, then determining which method is most appropriate in a given situation will be a matter of matching the study's aims to one or more appropriate methodologies. Mixed methods research can take several forms (Lancy, 1993, p. 11). In some cases, a quantitative study precedes a qualitative examination of the same question, although the reverse is probably more often the case. Some researchers in this tradition view qualitative methods as highly useful for exploratory research when little is known about the factors involved in a particular phenomenon (e.g., Creswell, 2009; Libarkin & Kurdziel, 2002; Marshall & Rossman, 2006). It is not unusual to find studies that are largely quantitative that have embedded within them a small qualitative piece (e.g., Dahl, Anderson, & Libarkin, 2005). Others view the two methods as complementary. A change from quantitative to qualitative methods or vice versa is warranted when previously published research on the subject has reached a point where a new perspective is needed (Hoepfl, 1997).

A comprehensive research programme designed to investigate how students understand a specific scientific concept will require a range of research strategies. In order to develop the fullest possible picture of how students understand that concept, we will need to collect and analyse data in a variety of ways. The picture will never be perfect for human beings are a complicated lot. A flexible use of methodologies over time is needed. Neither quantitative nor qualitative methods alone will suffice. At times that will mean a particular study will use only quantitative or qualitative methods. At other times, a mixed methods approach in the same study is called for. This often puts us in the best position to be able to offer sound pedagogical recommendations that practitioners will find useful.

Previous research on students' conceptions of deep time reviewed in chapter two employed a variety of methodologies. Studies and methods employed are delineated in Table 3.1. Classifying a particular study as quantitative, qualitative, or mixed methods is not always easy to do, especially if the researcher did not define the study in that manner. For the purpose of Table 3.1, the quantitative column includes studies that employed either descriptive or inferential statistics as the means to report and analyse data and draw conclusions. In a few cases, a few qualitative interviews were embedded within the larger quantitative design (e.g., Dahl, Anderson, & Libarkin, 2005; Dodick & Orion, 2003a; Marques & Thompson, 1997) but the study's design, analysis and reporting of data were mainly quantitative. The qualitative column includes studies that analysed data by searching for and reporting on themes that emerged. The mixed methods column can be especially problematic to define since it can include a variety of studies on a continuum from largely quantitative but with some qualitative analysis to the opposite scenario. Decisions about where to place a particular study were also based upon the authors' intent. One study that is listed in Table 3.1 as qualitative did employ some quantitative analysis (Libarkin et al., 2007). The authors describe their study as qualitative at several points, thus I placed it in that column. Table 3.1 demonstrates that quantitative methods have ruled the day in geoscience conceptions research on deep time. As I indicated previously, quantitative methods have many strengths and provide valuable data and interpretations. They do not, however, paint a complete picture. For that, we need additional methods.

Methodology	Frequency
Quantitative	14
Qualitative	3
Mixed Methods	1

Table 3.1 Frequency of methodology used in deep time studies reviewed in chapter two

3.2 Rationale for research methods for this study

I argued in chapter two that the geoscience research community needs a new lens through which to view how students understand deep time. It is also important that the present study fit into a larger methodological framework. In the previous section, I briefly described some of the strengths and weaknesses of quantitative and qualitative methods. I concluded the section with a description of the methods employed by others who have studied conceptions of deep time. There is another way to think about conceptions research that is important for understanding why I have chosen the methods for this study that I have.

Hashweh (1988) delineates three types of conceptions research studies: descriptive, explanatory, and intervention. Descriptive studies seek to explore the range of conceptions held by students, while the goal of explanatory studies is to determine *why* students hold the ideas they do. Explanatory studies can also provide information about why some erroneous conceptions persist while others are easily modified. As their name implies, intervention studies are designed to test the effectiveness of a particular teaching methodology or curriculum as a means to effect conceptual change. All three types of studies are needed, yet all three are not adequately represented in the deep time conceptions literature.

All of the studies dealing with deep time reviewed in chapter two can be described as descriptive. Taken together, they have demonstrated that students and adults at every age that was sampled hold similar, erroneous ideas about deep time. This is an important finding. Yet, we do not necessarily know *why* students are giving incorrect responses to our queries. Just as importantly, we don't know why some others are giving correct ones. It may well be that students are not confused about what we think they are, but are confused about something else entirely. Depending upon the response type chosen for a particular study, we may have overestimated some students' understanding, assuming that if they responded correctly they understood. That may not be the case. How then are we to interpret students' responses? What factors may influence how they comprehend deep time?

Review of previous studies on deep time point out two weaknesses in the existing research base on the topic. As was argued in chapter two from a theoretical standpoint, they have not looked systematically at the factors that underlie a concept of deep time. Thus, I proposed the model of the "three-legged stool." Here I make the argument that additional research is needed on methodological grounds. At this point an explanatory study is needed to address the fact that all research to date has been descriptive. Second, as Table 3.1 demonstrates, quantitative methods have predominated. Studies have primarily employed large-scale questionnaire type formats. Sometimes a few interviews were conducted as part of the larger study, but not always. While the large quantitative studies have been useful, additional studies employing qualitative methods in which students are asked to provide explanations for their responses are needed. Applying similar methodologies to those used in the quantitative studies can help determine whether the "three-legged stool" is a viable model. The themes that emerge from data analysis can provide an essential first step in a research programme to explore the utility of the model of the deep time "stool." This view reflects a pragmatist approach.

3.2.1 Why task-based semi-structured interviews?

Data collection in qualitative research can take many forms: observations, interviews, or the analysis of documents and artefacts (Creswell, 2009, pp. 179-180). Because this study is designed to investigate factors that may influence how students think about deep time, examination and analysis of documents or artefacts alone would not provide much useful information about how students were thinking about the subject. Observations of students as they work on tasks related to deep time are important. In fact, the tasks used in this study are a key component of the research design, but by themselves observations would provide insufficient data about students' thinking processes.

Posner and Gertzog (1982) describe interviews as a way to provide a window into factors that underlie specific conceptions and the relationships among those underlying factors. Interviewing allows the researcher to probe the thinking behind an individual's responses. Participants can be asked to clarify their responses and follow-up questions can be used to investigate particular avenues that may have been unanticipated prior to the study, giving them an advantage over paper and pencil surveys (Mahoney, n.d., Interviews section, ¶1). The interviewer can make counter-suggestions or simply even repeat what the interviewee has said in an attempt to elicit further information. The addition of specific tasks to the interview protocol permits the researcher to focus on students' strategies rather than whether answers are merely correct or incorrect (Goldin, 2000).

Task-based interviews were used extensively by Piaget (see section 2.2.1). Those methods and Piaget's highly detailed descriptions of his work (e.g., Piaget, 1969) have permitted other researchers to use similar tasks to independently verify and refute aspects of his theories. That is an important reason why interviews were chosen for this study. A number of the tasks in the interview protocol are modelled after ones used by others whose work was reviewed in chapter two. The similarity of tasks allows for comparison with previous studies and perhaps new interpretations of their findings. Besides the more general strengths of qualitative research described previously, task-based interviews can be used to investigate the strength of responses by looking for consistency across tasks. The search for common language across participants can be used to develop a theoretical model for how students acquire a notion of deep time.

3.2.2 Issues in the use of interviews

In a previous section I discussed some of the limitations of qualitative methods in general. All those apply to interviews as a qualitative data collection method. In this section I focus on weaknesses that are more specific to interviews.

The use of semi-structured interviews, task-based or otherwise, is not without concerns. Anyone employing this method of data collection must be aware of its limitations as well as its strengths. Interviews are heavily dependent upon the cooperation of the interviewees (Marshall & Rossman, 2006). Interviewee cooperation might be seen as a particular problem in other types of qualitative studies such as ethnographic or phenomenological research, but it is also a potential problem in a study of this type. Another is that the setting of the interviews is not always constant from interviewee to interviewee (Mahoney, n.d.), thus it is not always possible to control for background distractors. This can be a problem in a school setting where distractors are abundant. The interview itself likely has an effect upon students' responses. When individuals are repeatedly being asked to justify their responses, they may approach tasks in ways that are different from what they would if no one was asking them why they answered the way they did (Goldin, 2000). This may not actually be a limitation, but it certainly adds to the artificiality of the experience.

Researchers have been aware for some time that what interviewees say during the study may not accurately reflect what they truly think. Piaget, who used interviewing extensively, noted that children's responses may not be indicative of their actual conceptions (in Posner & Gertzog, 1982). In some instances the student is not particularly interested in the question and says the first thing that comes to mind, which is not necessarily what the person would say if asked the same question at another time. At other times, the way a question is asked leads children to particular responses, a point also made by others (Marin, 2004). Occasionally, children give the answer they think they are supposed to give. This can be a particular problem when interviews are conducted in a school setting. Some students may feel that the tasks require "school" answers, which may not correspond to what they would say if the interview was held in another place (Goldin, 2000). Even when students are giving thought to the question their responses may suggest a level of coherence and forethought that is not actually there (Myers & Newman, 2007). If asked similar questions at a later moment, they might reason quite differently. Further, in asking students to describe their thinking processes we assume they possess metacognitive awareness, when they may not.

The skill of the interviewer in both data collection and analysis has an important effect upon a study's outcomes (Marshall & Rossman, 2006). It was already mentioned that the wording of questions can lead participants to particular responses. An even greater problem is that of the perspectives that the interviewer and interviewee each bring to the task.

Johnson and Gott (1996) argue that researchers need to pay careful attention to methodological issues in the use of interviews and recognize the effect those issues have upon a study's outcomes. They are concerned that much science conceptions literature leaves one with the impression that uncovering students' ideas is "relatively straightforward" (p. 562). In reality there are several points where the process can break down. No matter what the subject of the interview, both the interviewer and interviewee bring their frames of reference to the task. The interviewer is often a domain expert and, hence, comes to the interview with domain vocabulary and knowledge. The interviewee will likely possess limited knowledge of the topic at hand, but will have some everyday knowledge. The interviewee's lexicon may contain some of the same words as those used by the interviewer but the definitions may differ. This can lead to a disconnect between what is said and what is perceived. Thus, it is possible that the interviewee will interpret the question being asked in a manner that is different from the one the interviewer intended. It is equally possible (and not mutually exclusive with the prior point), that the interviewer will interpret the interviewee's response in a manner that is different from what was meant (see also Myers & Newman, 2007). Thus, triangulation of responses (see section 3.1) is important if we are to have confidence in the credibility of students' responses.

Transcribing audio or videotapes can be challenging. While punctuation is used regularly in written communication it is not used in the same way when speaking. In a desire to transcribe in a manner that makes sense to a reader, nuances can be obscured. Unless videotapes are generated, nonverbal visual information is lost if the researcher does not record it. Interviews are also subject to potential biases on the part of the experimenter. Decisions about how data should be coded undoubtedly reflect a researcher's underlying assumptions. If coding categories were different, conclusions might also be different. The task-based interviews designed for this study may not tell us the relative importance of each of the three factors in an understanding of deep time.

All these concerns are inherent to the methodology, and minimally, call for caution in the interpretation of results. Task-based semi-structured interviews can, however, provide a place to begin a research programme. Themes that emerge from these interviews can be tested more fully using other methods. Designing an interview protocol that tries to uncover students' conceptions (and not simply the first thought that comes into their heads) is important for anyone using them as a research methodology.

3.2.3 Issues that must be addressed in the development of task-based interviews

One problem with interviews that was mentioned in the previous section is the possibility of misunderstanding between interviewer and interviewee. That problem could lead one to conclude that conceptions research is futile, but this is not so. Johnson and Gott (1996) call for conceptions researchers to do everything possible to place the interface between researcher and student on what they term "neutral ground" (p. 565). To place the interview situation on neutral ground means that,

> *The child understands what the researcher is asking in the meaning intended by the researcher, and the researcher understands the child's response in the meaning intended by the child. The neutral ground will be a much more limited affair than*

either the researcher's or the child's own frame of reference (Johnson & Gott, 1996, p. 565).

The authors are well aware of the difficulty of the task. They offer several guidelines that are useful when designing qualitative interview protocols. First, as much as possible it is important to construct items that neither lead participants in a particular direction nor limit their responses. The problem of different lexicons has already been mentioned. Researchers cannot assume that an interviewee holds a scientific conception simply because a scientific term is used. By the same token, using scientific terminology in interview questions can be problematic since there are no guarantees that the interviewee has the same understanding of the term as the interviewer. Johnson and Gott call for the triangulation of responses, a point that was made earlier.

3.2.4 Why middle school, high school, and university students?

The National Science Education Standards described in section 1.3 show that geoscience topics that relate to deep time appear in the U.S. curriculum most commonly at the middle school (ages 11-14) and high school levels (ages 14-18). In order to make sense of things such as tectonic plate motion and the fossil record students will need to possess some understanding of deep time. It makes sense to include students from both middle school and high school in the sample for this study.

The geoscience content taught in elementary school (ages 5-11) focuses primarily on Earth materials rather than Earth processes, thus, children at those ages are not included in this study. (The water cycle is an exception that is often taught in elementary school and then retaught in either middle or high school.) Therefore, knowledge of deep time is not so crucial to understanding geoscience content at the elementary level. Further, as was seen in chapter two, elementary school age children have sufficient difficulty with longer units of conventional time. Deep time would be expected to be quite problematic for them.

Many U.S. universities have introductory geology courses that students across the university take to fulfill general education requirements. Thus, they contain students who are science majors and those who are not. For non-science majors this will often be the final science course ever taken in their educational careers. Those introductory courses deal with a variety of Earth processes that occur in the context of deep time. Research with this population has demonstrated that they have a rather poor understanding of the topic near the end of the course not just at the beginning (Libarkin & S. Anderson, 2005). The fact that their understanding of deep time does not improve significantly by the end of the course makes them an important group to include in this sample. A study designed to explore the factors that influence an understanding of deep time may be a first step in explaining why university students still struggle with the concept at the end of a geoscience course.

3.2.5 Why a cross-age study?

Once the decision to focus on middle school, high school, and university students was made, the next decision was whether to conduct a longitudinal or a cross-age study. A cross-age study was chosen because it fits the study's aims while a longitudinal study does not. A cross-age study does not permit a researcher to investigate the development of a particular individual's understanding. However, this investigation is not designed to uncover the process by which individual students come to understand deep time. Nor is this study designed to invoke a change in a student's conceptions. It does, however, allow one to explore how age may be related to student responses. While overall understanding of the concept doesn't appear to change much with age, the reasons for why students hold those conceptions might vary with age. Perhaps one of the three factors being explored will be a greater factor in a student's difficulties with deep time at different ages. This type of methodology is also likely to uncover a range of responses that can form a basis for later follow-up research. A cross-age study appears to be the best design for the research questions explored in this particular investigation.

3.3 Development of the instrument: The role of preliminary interviews

A series of preliminary interviews based upon the literature reviewed in chapter two preceded the interviews for the main study. Results of those early interviews were crucial in the design of the final interview protocol. Hence, they are briefly described here. There was a synergy between the preliminary interviews, their analysis, and the literature review. Responses in those early interviews suggested further avenues to explore in the literature. Thus the base of the literature review was broadened considerably to include literature in each of the three areas outside a deep time context. At the same time, additional reading of the literature indicated ways in which the interviews could be modified. For example, errors commonly made on conventional time tasks of duration (see section 2.2.3) such as the equating of spatial distance with duration pointed to ways to modify the interview protocol. Results of the preliminary interviews worked in tandem with the literature review described in chapter two and resulted in the development of the model of the "three-legged stool."

The purpose of the preliminary interviews was two-fold. First, the items in these interviews were designed to investigate what types of questions should be asked in the main study. While some of the items were common to all exploratory interviews, the questions evolved as a result of analysis of interview transcripts. In that sense, there is an emergent quality to this study, although ultimately the interview protocol in the actual study did not vary from individual to individual. Second, the exploratory interviews provided an opportunity for the investigator to practice the interview process before beginning the actual research study.

Preliminary interviews were conducted in three stages. All participants in the preliminary interviews comprised a convenience sample. First, four university undergraduates (three female and one male) were interviewed. The interview protocol was subsequently revised based upon analysis of their responses. One 12-year-old female and two 14-year-old males (middle school students) were then interviewed. Following analysis of these responses, the interview protocol was again revised. This revised interview protocol was administered to two university undergraduates. All participants in the preliminary stage were known to the investigator. All interviews were audio recorded. Three university student interviews and all middle school interviews were fully transcribed for analysis.

Some of the questions in the preliminary interviews were set within the context of deep time while others were well within conventional time. There were fewer tasks in the initial interviews. One item had nothing to do with deep time at all. Students in those very first interviews:

- Sequentially ordered cards of nine geoscience events from the Big Bang to the extinction of woolly mammoths, a task modelled after Trend (2000) but with the inclusion of a card, “Great Pyramids of Egypt were built”
- Drew a cross-section of the Earth as if it had been cut in half from North to South
- Placed events in a day on one timeline and major holidays in a year on a separate timeline
- Placed four events from Earth’s history on a timeline, taken from the Geoscience Concept Inventory, question 28 (Libarkin & Anderson, 2005)
- Watched four animations, all of which are available online: the break-up of Pangaea, an oceanic-oceanic divergent plate boundary, an oceanic-continental divergent boundary, and a continental-continental convergent boundary and answered questions about what they thought was happening and how much time was required for the events in the animations to happen

A few items were changed prior to the interviews with the younger pupils. The one-day and one-year duration timelines were removed after analysis of the interview transcripts with the first four university students as they did not appear to yield any useful information. An additional item was added for the younger pupils based upon work done with size of scale phenomena (Jones et al., 2009; Jones et al., 2008; Tretter et al., 2006; Tretter et al., 2006). This item contained a list of events that ranged from very short (amount of time to eat dinner) to very long (amount of time for Pangaea to break apart). Students were asked to choose the duration of each event from among time categories that ranged from seconds to millions of years. This item was added for two reasons. The first was to compare students' responses on this item and others in the interview. The second was to compare students' ability to estimate durations for events within a human lifetime and those outside it.

Several themes emerged from participant responses across all the preliminary interviews. These themes were used to revise the instrument for the actual study. One commonality among responses of both university and middle school students was an inability to work with numbers greater than thousands. When talking about durations of events, only one of the interviewees used numbers greater than thousands of years. This finding served to corroborate what others researching deep time conceptions had already said about the issue of large numbers (Oversby, 1996; Trend, 1998). Thus, additional items were constructed to specifically probe how students understand large numbers both in and outside a geoscience context.

If asked to place the duration of events in broad time categories similar to what Trend (Trend, 2000, 2001b) did for succession, students were able to place durations of geologic events in those categories, but when asked to offer their own judgment about how long a geologic process took, with only one exception (discussed below), interviewees never mentioned a time frame longer than a few thousand years. One university student's response was typical. When asked how long it took for Pangaea to break apart she replied, *I'm pretty sure that would take a really long time because it has a far way to go. The plates don't move very fast. I would say a thousand years or something.*

A 20-year-old male university student (history major) was the outlier in the group. This young man has a self-described aversion to mathematics and science. He has had no higher level coursework at university dealing with geoscience, and it is not a subject about which he reads outside of school. During the interview he commented that there was really no difference in 1,000 years or one year in terms of geologic time—they were the same. This understanding was not demonstrated by any of the other interviewees—middle school pupils or university students. In the context of geologic time, their spontaneous comments indicated a view that a year was much longer than a day. Within a human lifetime that is true, but within the context of deep time the differences are inconsequential. This young man also indicated that he would use a historical event (the American Revolution) as a unit with which to determine how long ago a particular event occurred. He said that since the American Revolution happened a little over 200 years ago he would ask himself how many 200-year periods have happened since the event in question occurred. In contrast, one of the middle school pupils illustrated that it was possible to have a bit of information but be unable to use it. He said he had heard from his sister that the Earth was five billion years old; yet, he was unable to use that information to reason about ages and durations. The number had little meaning to him.

Several students appeared to equate size of a feature with the amount of time necessary for it to form. For example, one student said that it would take much longer for the Colorado River to carve the Grand Canyon than for a coral reef to form because coral reefs are smaller than the Grand Canyon. In his view, “It takes a long time to make a giant hole in the ground.” This suggested some inability to apply a conventional understanding of time in a geologic context. In other words, students made a similar error to one described in chapter two in which distance was equated with duration. Here, amount of space (size) = duration.

In explanations for their answers participants said they were unsure about some events. Two of the three middle school pupils had not heard of the Big Bang. As a result, they were unable to correctly place the Big Bang in a card sequence of geologic events. This was in agreement with findings reported in chapter two and suggested that lack of geoscience content knowledge may play a significant role in student responses.

Additionally, responses across tasks were sometimes inconsistent. This underscored the need for triangulation of responses that Johnson and Gott (1996) indicate is critical if findings are to be trusted. Further, a participant’s belief system influenced responses. One person felt she knew what the scientifically correct response should have been but rejected it as inconsistent with her religious beliefs. The young woman’s views were an example of factors that could influence the stability of the deep time “stool” that were discussed in section 2.8. This pointed out the need to both recognize the role of beliefs in responses and to choose a sample that would represent that demographic but not overly so. A fuller discussion of this point follows in 3.5.1.1.

Following the middle school interviews, the protocol was revised again to include all the items that were a part of the final interview protocol. These changes were made to more fully explore each area of the "stool" within a geoscience context and outside it. Several items (all of which are discussed below) were added. Specifically, items that adapted methods used by others (Ault, 1980; Dodick & Orion, 2003a) were included in order to facilitate comparison with their findings. Some items were revised to make their meaning clearer. For example, participants seemed confused when asked to indicate the amount of time necessary for "the start and finish of an earthquake." Responses indicated that some individuals were focusing on the stress leading up to the actual event rather than the duration of the vibrations during the earthquake. This item was reworded to "the amount of time the ground shakes during an earthquake."

This revised interview protocol was administered to two female university preservice elementary teachers (primary teacher trainees). The two women were asked to provide feedback as to the clarity of the interview protocol. Audiotapes were also reviewed to check for item clarity and whether or not questions appeared to be leading students to particular responses. No further revisions were made to the interview items after these two students were interviewed.

3.4 The interview protocol

The final decision for what items to include in the interviews flowed out of the literature review in chapter two and the themes that emerged from the preliminary interviews discussed above. Items were designed to explore each of the three "legs" of the deep time "stool." Some items can be triangulated with other items in the interviews to ascertain consistency of responses. Figure 3.1 is a schematic of how the interview tasks fit together. Succession and duration were the organising principles around which the tasks were structured. Duration items outnumbered succession ones since research dealing with how students understand duration is underrepresented in the literature. Each item looked at succession or duration in the context of one of the three “legs” of the “stool”: conventional time, large numbers, or geoscience content knowledge. Items that addressed each of those categories are shown in the centre. A number of the items provide information on several categories simultaneously. Items are described briefly in the next few sections. A complete transcript of the final interview script along with correct answers to the items can be found in Appendix A.

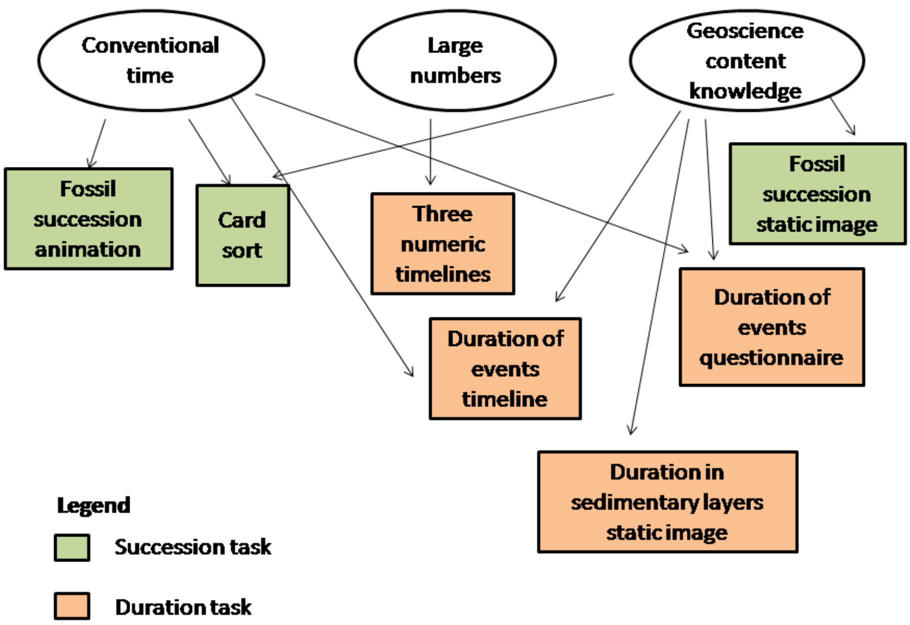


Figure 3.1 Relationships among items in the interview protocol

3.4.1 Succession items

Succession in time involves the before and after relationship as well as the ability to discern that two events occur simultaneously. One succession item was adapted from Puzzle 5 of the GeoTAT (Dodick & Orion, 2003a, 2003b). Puzzle 5 can be found in Figure 2.2 (p. 126). In the GeoTAT students viewed Puzzle 5 and were asked to place the fossil layers in temporal order by appearance in the columns. In the present study, students were shown the image and asked to indicate whether the trilobite or the brachiopod was formed first (trilobite is correct). This pair was chosen because both appear at the bottom of a column, but a brachiopod also appears higher up in the second column. This question should be sensitive to prior knowledge. A student would need to know that, unless deformation of strata has occurred, older layers in sedimentary strata are found below younger layers (superposition). However, that inference may not require any particular *geoscience* knowledge as it conforms to an everyday idea. When I see a brick wall, I know that the layer of bricks on the bottom was put there before the layers above them. The problem is that the brachiopod and the trilobite are both on the bottom of a column. Thus, I must also realize that two brachiopods that appear in different places represent the same depositional episode. This does require geoscience content knowledge.

An animation was created from Puzzle 5 so that students could watch the fossil layers appear in real time. This was done to enable the task to be completed based solely upon an understanding of succession without the requirement for the geoscience content knowledge described in the previous paragraph. A still image of the animation based upon Puzzle 5 is shown in Figure 3.2. In the animation each layer appeared on the screen as would be appropriate based upon fossil succession. Participants watched the animation on a laptop which they were able to move so that it was a comfortable distance for viewing. They were told they could watch it multiple times both before and after being asked questions about it. Students were then asked to compare three pairs of layers and say which one formed first or if they formed at the same time. The pairs of layers were:

- Clam vs. fish scale: chosen because both appear at the same place in adjacent columns, while the clam also appears at another position in the first column (clam is correct)
- Coral vs. brachiopod: chosen because the brachiopod in the second column occupies a higher position than the coral in the first column while the brachiopod in the third column occupies a lower position (coral is correct)
- Trilobite vs. brachiopod: chosen to serve as a comparison to a student's previous response to this question when it was asked prior to watching the animation (trilobite is correct)

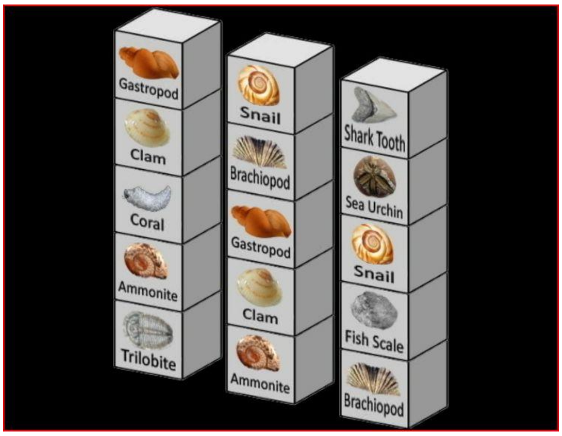


Figure 3.2 Animation exploring ability to apply succession in conventional time

Finally, they were asked to sequence the layers in the three columns, as Dodick and Orion did with the static image. This requires the before and after relationship since I must understand that if I saw the trilobite appear on the screen before the ammonite what I saw represents temporal succession. I must also be able to judge that if two layers appear at the same time [ammonite in first column and ammonite in second column], they represent simultaneous events. These last two statements are undoubtedly self-evident to the reader, however, that is not the case for individuals who do not possess a solid concept of succession in conventional time (see section 2.2.1).

An additional set of items explored succession and was modelled after Trend's work (1998, 2000, 2001a, 2001b) discussed in section 2.6.1. A complete list of items in correct order can be found in Appendix A and also Table 4.12 on p. 278 under the column Rank: Consensus. The difference between this task and Trend is that all his events occurred in deep time with the exception of one named "present day." Nine events in deep time were used for the present study. An additional four events were added from ancient history through the Age of Exploration. Adding these items provided a means to test understanding of succession in both conventional and deep time simultaneously. In U.S. schools, ancient history and the Age of Exploration are frequently taught prior to eighth grade and again in high school. Therefore, all participants in the sample should have had some exposure to these historical events. If students find it no easier to place events considered part of recorded history than those in deep time it could suggest that specific knowledge of the events may be a more important factor in their responses than how long ago the events occurred. In fact, one might argue that sequencing historical events accurately might actually be more difficult than sequencing at least some of the geoscience events, precisely because they are more sensitive to subject matter knowledge. It might be possible to sequence several of the geoscience events using logical reasoning without specific geoscience content knowledge.

Following the initial card sort, participants were asked to place cards into piles based upon how long ago the events on the cards occurred. They were not given the categories but were told to determine the appropriate number of categories themselves. This is similar to a task described in chapter two with scale related to size (Tretter et al., 2006a; Tretter et al., 2006b). This method allowed participants to organize the categories themselves and could provide some insight into how students view the relationship of the events to each other. Students were then asked to provide a name for each category. If they did not mention an age, they were asked to provide an age for their categories.

3.4.2 Duration items

Three computer animations were created to address how students perceive duration in conventional time. They were modelled after the types of tasks described in *The Child's Conception of Time* (Piaget, 1969) which were reviewed in section 2.2.1. Students watched these animations on the same laptop used for the fossil layers animation described earlier. Still images of the animations appear in Figure 3.3, 3.4, and 3.5. All have the same horizontal coloured layers that appear against a black background in the following order from bottom to top: red, blue, yellow, green, brown, and pink. Each layer fills up one after the other. Animations run for between

33-41 seconds. There is a timer that runs in the upper right portion of the screen although participants were not specifically directed to it. Table 3.2 lists the layers for each animation and the amount of time it takes for each layer to fill. In the first animation all six layers are of the same thickness, but fill at different rates. Duration (amount of time to fill) varies but total work done (in a Piagetian sense) is the same. In animation two, the layers are of varying thicknesses but all take six seconds to fill. Duration remains the same but amount of work done (thickness of layers) varies. In animation three, both the thickness of the layers and the time required for them to fill are varied.

Figure 3.3 Animation 1

Figure 3.4 Animation 2

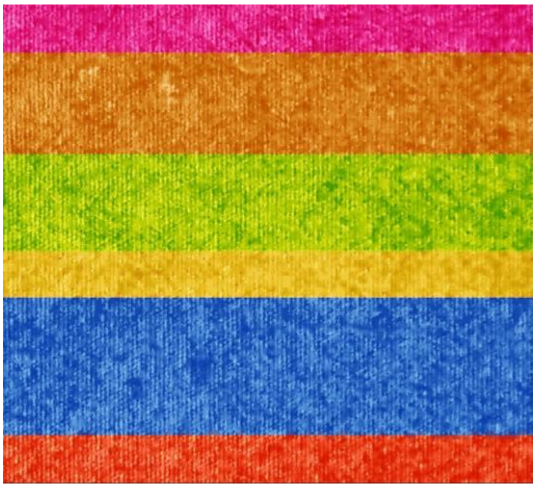


Figure 3.5 Animation 3

Colour (from bottom to top)	Animation 1	Animation 2	Animation 3
Red	5 seconds	6 seconds	6 seconds
Blue	8 seconds	6 seconds	4 seconds
Yellow	4 seconds	6 seconds	7 seconds
Green	6 seconds	6 seconds	7 seconds
Brown	4 seconds	6 seconds	10 seconds
Pink	6 seconds	6 seconds	7 seconds

Table 3.2 Amount of time to fill coloured layers in each animation

Research post-Piaget described in chapter two noted the strong relationship between temporal and spatial mapping for individuals well beyond the ages of the children used in Piaget's research (Boroditsky, 2000; Casasanto & Boroditsky, 2008; Friedman, 2005). The question was whether spatial cues (thickness of the layers) would influence judgments of duration. The animations are *not* designed to simulate actual deposition. The coloured layers appear smoothly in the animations whereas deposition is much more uneven. They are analogous to deposition only in so far as it occurs in horizontal layers that exhibit superposition under normal circumstances.

Rather, the purpose of the animations was to take horizontal layers and determine if students were able to correctly judge duration in conventional time even when spatial and temporal information did not co-vary (i.e., a thinner layer filled more quickly than a thicker one or a thicker and thinner layer filled in the same amount of time).

After watching each animation participants were asked two questions that required them to compare how long it took two specific layers to fill. They were told they could watch each animation as many times as they wished both before and after the interviewer asked questions. This was so participants could focus on the perceptual task and not need to rely on memory to answer the questions. Half the participants viewed animation one first and half watched animation two first. It was not known whether order of presentation would affect student responses, but this was investigated in the data analysis.

Table 3.3 lists the specific layers for which students compared durations on each animation. Correct answers for each pair are in bold. Neither answer is bolded for questions following Animation 2 since all layers required the same amount of time to fill. These specific pairs of layers were chosen based upon the literature reviewed in chapter two and the type of errors described there on duration tasks. For each animation the first question asked, “Which layer took longer?” The second question asked, “Which layer filled more quickly?” Comparison layers were chosen so that for question one, the first layer of the pair seen might elicit the type of error seen in the literature. The second layer seen was more likely to elicit an error like those reported in chapter two for question two.

Animation	Layers compared
Animation 1 (A1)	Red vs. **blue**
	Yellow vs. green
	Blue vs. yellow (same)
Animation 2 (A2)	Brown vs. pink (same)
	Green vs. **brown**
Animation 3 (A3)	Red vs. **blue**

Table 3.3 Coloured layers compared on duration animations

The animations required different strategies to answer the questions. For A1, one only needed to take account of different rates since the thickness of the layers was held constant. In A2 one had to account for both the thickness of the layer and the rate at which it filled to determine duration. A3 included one of the A1 type and one question unique to the animations. Rate and size varied inversely for this question; however, the rate difference between the two layers was greater than for any other pairs compared across animations. Alternatively, durations for all animations could be judged by counting the time it took each layer to fill, either in one's head or by using the timer that was on the screen.

The next duration item occurred immediately after the animations. Participants were shown a line drawing of a sedimentary stratigraphic sequence (Figure 3.6) typical of what one might find in the Southwest United States. They were asked to use information from the animations to compare layers 3 and 4 and determine which one probably took longer to form. Students were also told that neither the designs on the layers nor the erosional patterns had anything to do with how long layers took to form. Since no information is provided about the depositional environment, the correct answer to this question is that it is impossible to say for certain which layer took longer to form based upon the drawing alone. This is consistent with the animations as there was no consistent pattern across animations in terms of thickness of the layer and time required to fill. Ault (1982) alleged that elementary school students had no difficulty understanding conventional time but were unable to apply it to a geologic context. This item was designed to explore whether or not that is so.

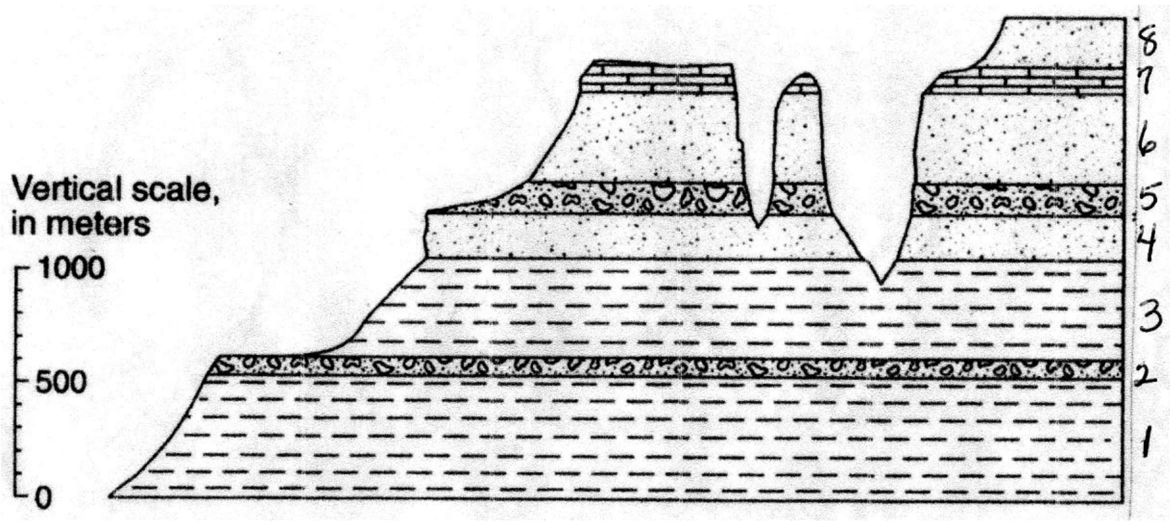


Figure 3.6 Line drawing of a hypothetical stratigraphic sequence

The next set of duration items were developed as a result of the preliminary interviews described in section 3.3. As was noted earlier, participants in those early interviews had some difficulty conceiving of large numbers consistent with the literature on the subject reported in chapter two. Additionally, chapter two pointed to the lack of research into how people conceive of durations over long periods of time. Some research has explored how people perceive how much time has passed from the occurrence of an event to the present time, but not how they understand the length of the time period represented by a specific unit of time. This set of questions was designed to address this issue.

Students were asked to create four timelines, each 50 cm long. Three of them only involved numbers, albeit of various magnitudes. The other line involved both large numbers and geoscience content knowledge. For the timelines dealing only with numbers, participants were directed to place times on the timeline based upon how long they take "in proportion to each other." See Table 3.4 on p. 181 for an explanation of items in each timeline. Appendix A includes possible correct timelines. In each case the order of items in Table 3.4 corresponds to the order in which items were presented. For example, 1 day, 1 month, 1 year, and 100 years were written on four separate index cards. Cards were placed in front of participants in left to right order with 1 day farthest left and 100 years farthest right. If students expressed hesitancy or confusion about the directions, they were shown a timeline made by the investigator for one minute, one hour, and one day. Younger students were more likely to be shown the investigator's timeline than older participants. Successful completion of Timeline 1 required knowledge of the proportional relationships between a day and a month and a month and a year, thus, some knowledge of units of time was required. The proportional relationship between one year and 100 years might be determined purely on the basis of numerical reasoning since the unit (year) was the same. Timelines 2 and 4 could be completed by using proportional reasoning since, again, the unit (year) was the same.

Timelines were compared to determine if there was a difference in strategy or placement of times on the timelines. In other words, would students find it easier to proportionally place the times when there were fewer of them and when the numbers were smaller? Times on Timeline 1 were chosen for their familiarity and the fact that three of the four times have been experienced by all participants in the study. The fourth is often viewed as a long time in terms of human experience. Times for Timeline 2 were chosen for several reasons. First, they contained some time periods that fall within recorded history and others that do not. Second, all relate to each other by powers of ten. However, they do not all relate in the same manner (not all multiplied by 100, for example). While, students might be expected to find it easy to relate 1,000 to 100,000, relating 100,000 to 1 million might prove more challenging.

Timeline 1	Timeline 2	Timeline 3	Timeline 4
1 day	1,000 years	Earth spinning around once	1 minute
1 month	100,000 years	How long most coral reefs have been growing	1 day
1 year	1 million years	Break-up of supercontinent Pangaea	1 month
100 years	100 million years	Earth going around Sun once	1 year
		Moon going around Earth once	10,000 years
		Carving of Grand Canyon by Colorado River	10 million years
		Time ground shakes during an earthquake	100 million years

Table 3.4 Items in Timelines 1-4

On Timeline 3 participants were asked to place the events above on a timeline based upon how long they take to happen relative to each other. This is a variation on the first two timelines where students simply placed numbers on the line. For Timeline 3 they had to first determine how long each of the events takes to occur and then compare them to one another to decide on their relative placement on the timeline. Thus, this item required knowledge of the duration of specific events (geoscience content knowledge) in addition to an understanding of large numbers.

In order to get the numbers for Timeline 4, scientifically accepted times for each of the events on Timeline 3 were rounded to the nearest power of ten or nearest familiar unit of time. These numbers appear in column 4 of Table 3.4. Participants were asked to place the numbers for Timeline 4 on a line in the same way they did for Timelines 1 and 2. This final timeline was added for two reasons. While it is similar to Timelines 1 and 2, it combines times both within and outside individual human experience. Times outside individual lived experience encompassed times which are part of human history and those that are outside human history. The second reason was to see if students approached this timeline differently after completing Timeline 3. The decision to round times to the nearest power of 10 was made because it was felt that using other numbers would add additional complexity to the task that would not yield significant information. In terms of deep time, there is no appreciable difference between 10,000 and 20,000 years.

Deep time encompasses time periods in the billions, but the largest time period in the timelines was 120 million years, or the amount of time for Pangaea to break apart. This number was rounded to 100 million for reasons that were described above. Billions were purposely not included. Timelines 3 and 4 involved numbers that differed by more than eight orders of magnitude. It was felt that the addition of a number in the billions would not provide significantly more information than 100 million. There was nothing in the literature reviewed in chapter two that would suggest that an additional order of magnitude would make any difference in the nature of students’ responses when dealing with numbers at this scale.

One duration item in the interview could most accurately be viewed as a geoscience content knowledge item. Students were given a list of 20 items and asked to indicate how long each takes to occur (see Appendix A). This is modelled after a similar task with size used by Tretter, et al., (2006a; 2006b). Choices were seconds, minutes, days, years, hundreds of years, thousands of years, and millions of years. Some time periods were deliberately not added as choices because of the potential ambiguity. For example, one could reason that a pumpkin seed grows into a fully ripe pumpkin in days, weeks, or months. To eliminate this potential source of confusion weeks and months were not included as choices. Since some of the events in this item were also in Timeline 3, durations for the two items can be triangulated. The responses of someone who indicated a different duration for an event in this item than what was indicated for Timeline 3 may not be valid. Other items were chosen as ones with which students would likely be familiar, such as the amount of time necessary to eat dinner or drive from one side of the state of Pennsylvania to the other (the state where participants lived at the time of the study). A final group of items was chosen because their durations were likely to be unfamiliar to students. These included events with short durations such as the amount of time necessary for a red blood cell to travel through the entire body as well as longer durations such as the amount of time necessary for the Colorado River to carve the Grand Canyon. The purpose of these items was to explore the question of whether subject matter knowledge of any type is a factor in judging duration irrespective of its length.

3.4.3 The interview procedure

All interviews were conducted at the school or university at which the student was enrolled in a room with a small table and chairs and were audio recorded for later transcription and analysis. The interview protocol was approved by the Ethics Advisory Committee at Durham University as meeting all requirements for the ethical treatment of human subjects. A sample consent form signed by university participants and one signed by parents of eighth and eleventh graders are in Appendix B.

A few minutes at the beginning of the interview were spent getting to know students, asking them about themselves, and their previous exposure to geoscience topics. After each task, participants were asked to explain their thinking or why they completed the task the way they did. Students were told at the beginning that they could ask questions at any time during the interview. They were told they could think out loud as they worked if they would like. They were encouraged to give responses throughout the interview but were told they didn't have to guess. If they weren't sure they could simply say they didn't know. Follow-up questions were asked as appropriate. Students were given the opportunity to change their mind about their answers. When there were discrepancies in responses they were pointed out to participants. They could then elect to change their answers or clarify them. Although, they were not specifically told they could take notes if they wished, several students used the paper and pencil that were on the table to do so.

3.5 The sample

The sample consisted of 35 participants: 12 university undergraduates, 12 eighth graders and 11 eleventh graders. Interviews were conducted in fall 2007 and spring 2008. This sample could be described as purposive for several reasons. First, an attempt was made to balance participants by gender, although this was not successful with the 11th graders, a point that will be discussed later. Second, science teachers of the 8th and 11th graders recommended students who they identified as high, middle, or low achievers in science. This was done to include pupils who represented the range of ability levels. Finally, one-half of the university students were selected from a group that is frequently deemed to hold metaphysical objections to the concept of deep time. This is discussed in section 3.5.1.1. No attempt was made to choose a sample that represented the racial and ethnic diversity found in the particular school or universities from which this sample was obtained. Students were not chosen to represent the range of socioeconomic levels or their proportions within the larger student body of their school or university. All university students self-selected to participate. They are unlikely to represent the ability levels or range of majors in their university.

Information is provided about each subgroup within the sample, along with relevant information about the types of instruction these students are likely to have received prior to the interviews. Information regarding topics participants studied may be useful to the reader. However, as some research in chapter two demonstrated (e.g., Catley & Novick, 2009) the extent of prior coursework regarding deep time does not appear to have an appreciable effect upon conceptions of those who can best be described as novices. Additionally, this study is designed to investigate factors that may underlie students' conceptions not explore the relationship of prior coursework to those conceptions.

3.5.1 University Participants

University participants were from two institutions. A total of five males and seven females participated and represented a variety of majors offered by their institutions. All participants were volunteers. Students ranged in age from eighteen to twenty-four with a median age of twenty. A list of university participants with their ages and majors can be found in Appendix D.

3.5.1.1 Institution A

The first six attend a small religiously-affiliated liberal arts college in eastern Pennsylvania, United States in which approximately half of the students are religious vocation majors (Institution A). Undergraduate enrolment at the time of the interviews was 944. The school population is 51% female and 49% male with 81% of the student body identifying themselves as white, non-Hispanic. The Scholastic Aptitude Test (SAT) is an exam required by many U.S. universities for admission. While it is not required by Institution A, an annual ranking guide to U.S. colleges and universities indicates that 25% of admitted students who provided SAT scores to the institution scored below 860, and 25% of admitted students scored above 1180 (Zuckerman, 2009). This means that those students who took and reported their SAT scores to the college are in the average range of freshmen admitted to university in the U.S. However, since the test is not required for admission to Institution A it cannot be assumed that the student body as a whole falls into the average range. In the U.S. the percentage of students receiving financial assistance is an approximate indicator of the overall socioeconomic level of the student body. Ninety-eight percent of the students at Institution A receive some sort of financial aid ("Search For Schools, Colleges and Libraries" n.d.). In 2003-2004, the national average was 63% for all colleges and universities in the U.S. ("Fast Facts," n.d.)

Three males and three females participated in the interviews. All were enrolled in an introductory Earth Science course taught by the investigator in the fall of 2007 and were interviewed approximately two-thirds of the way through the course (November 2007). The course is designed to meet general education requirements for a science course. In the U.S., introductory courses include students from a variety of majors and not simply those who are studying science. Two were preparing to be elementary (primary) teachers, one was a business major, and the remaining three were preparing for religious vocations. The investigator asked for volunteers to participate in the study at the start of one class session. Volunteers contacted the investigator individually to signal their willingness to participate. All six indicated some prior earth science instruction in high school, and this was the first university geoscience course for all students. Prior to the interviews, students in the course studied tectonic and surface processes that produce Earth's changing features. Rocks, minerals, and the rock cycle were covered early in the course. They completed a section on geologic time that included relative and absolute dating methods used in geoscience and had a brief introduction to the nebular hypothesis, though a fuller discussion of the origin of the solar system occurred after the interviews took place.

It might be argued that this portion of the sample is purely one of convenience, yet it is also purposive. Many students at Institution A come from backgrounds in which there is a perceived conflict between religious teachings and scientific understanding regarding the age of the Earth and the universe. Therefore, students at this institution may be more likely to demonstrate a philosophical unwillingness to acknowledge deep time than those in the larger U.S. university population. They come from an important demographic group for people researching understanding of deep time in the U.S. since they represent a view that is widely held within society at large. According to a 2007 Gallup survey, 43% of Americans believe, "God created human beings pretty much in their present form at one time within the last 10,000 years or so." ("Evolution, Creationism, Intelligent Design", n.d.) As was noted in chapter two, other researchers in the U.S. have found that some students appear to reject the notion of deep time based upon their metaphysical beliefs (e.g., Libarkin, et al, 2007). Therefore the inclusion of this group of individuals in the study was *apropos*. However, it would have been inappropriate to have selected the entire university sample from this population as the results may have been skewed thereby making the study's conclusions suspect.

Additionally, these six students were the only ones in the study known to the examiner prior to the interviews. The pre-existing teacher/student relationship makes the interview dynamic different with them than with any of the other participants. However, since Institution A is a small college, the examiner is the only instructor who teaches a geoscience course. Thus, it was not possible to interview students at this institution who were taught by someone else. The purpose of the study justifies this group's inclusion for the reasons stated above.

3.5.1.2 Institution B

An additional six students were interviewed from a comprehensive state university also located in eastern Pennsylvania (Institution B). This school's reported undergraduate enrolment at the time of the study was 10,818. The student body is 62% female and 38% male. Eighty-six percent of students identify themselves as white, non-Hispanic. The SAT is required for admission to Institution B. Twenty-five percent of admitted students scored below 970 and 25% scored above 1150 (Zuckerman, 2009). Like Institution A, students at Institution B are in the average range for SAT scores for incoming freshmen. However, that fact is deceptive since Institution B requires the SAT for admission while Institution A does not. Overall, students at both institutions fall within the average to below average range of students at all U.S. colleges and universities. Seventy-two percent of students at Institution B receive some sort of financial aid.

All students from Institution B were enrolled in an introductory geology course in spring 2008 and were interviewed approximately two-thirds of the way through the semester (March 2008). This is a course that includes both geology and non-geology students and is designed to meet general education requirements for a science course. Students came from several sections of the course taught by a variety of instructors, none of whom was involved in this study. Instructors of the respective sections made a general call for volunteers for the study during a regular class meeting. After initial attempts to obtain volunteers were unsuccessful, a second request was made. All introductory geology students received a flyer requesting their participation, and they were offered $15 for their assistance. Two geography majors, one history major, two computer science majors, and one geology major participated in the interviews. Five of the six indicated some prior earth science instruction in either middle school or high school, and this was the first university geoscience course for all participants including the person who said he was planning to major in geology. Instructors of the various sections of the course employed a similar course outline. Prior to the interviews, all students had studied plate tectonics, with attention to the tectonic processes responsible for mountain building. They had all been taught about rocks, minerals, and the rock cycle. These students had also completed a unit on geologic time with attention given to relative and absolute dating methods.

Not all students from Institutions A and B completed their compulsory education in Pennsylvania. High school graduation requirements vary somewhat from state to state within the U.S. though all would provide instruction on place value, proportional reasoning, and scientific notation. American universities require graduates to complete a mathematics course at the university level (or show proficiency), but that course can be taken at different times throughout a student's university career. Further, at both Institution A and Institution B, a student many choose from a variety of courses to meet the mathematics requirement. Thus, it is likely that university participants in this study vary considerably in their mathematical skills.

3.5.2 Eighth and eleventh grade participants

Eighth and eleventh grade participants attend a public charter school in eastern Pennsylvania. This charter school is a publicly funded school that is granted the right to operate independently of some of the state and local school district regulations to which other U.S. public schools are subject (i.e., analogous to a UK trust school). Students in many charter schools in Pennsylvania come from several neighbouring school districts including the district in which the school is situated. This particular school serves students in kindergarten through grade 12. Recent publicly available data (2007) indicates a student body of 911, 63% percent of which are white, non-Hispanic and 29% of which are African American. In the U.S., the percentage of students in a school who are eligible for free or reduced lunch is used as an indicator of the socioeconomic level of the population. Twelve percent of the students were eligible for free or reduced lunch in 2007 ("Schooldigger.com -- Search and compare elementary, middle, and high schools.," 2008). The state average for Pennsylvania in 2007 was 35% ("Food & Nutrition: National School Lunch Program," n.d.). Thus, the socioeconomic level of the student body in this school is above average compared to other schools within Pennsylvania. Eighth and eleventh graders were chosen by their science teacher to participate in the interviews. All were interviewed in spring 2008, approximately one month before the end of the school year. The teacher selected two males and two females from each grade who the teacher rated as high, middle, or low achievers. Six boys and six girls in eighth grade were interviewed. They ranged in age from 13 years, 9 months to 16 years, 1 month, with a median age of 14 years, 4 ½ months. Several of the eleventh graders chosen by the science teacher did not return the parental consent form for the interview and were therefore unable to participate. As a result of teacher substitutions to the original list, the total number of eleventh graders interviewed was eleven. In the end, there were one female and two male high achievers, two female and three male middle achievers, and one female and two male low achievers. The seven males and four females ranged in age from 16 years, 3 months to 17 years 11 months, with a median age of 17 years, 1 month. A list of the 8th and 11th grade participants with their ages and achievement levels as ranked by their science teacher can be found in Appendix E. All students in the 8th and 11th grade samples said they had studied earth science two years prior to the interviews in 6th grade and 9th grade, respectively.

Table 1.1 on p. 14 provides a framework for the types of topics 8th and 11th grade participants would have studied in previous earth science courses. Not all 8th graders could list specific topics they had discussed in earth science. Those who did so mentioned plate tectonics, mountains, volcanoes, and earthquakes. No one mentioned fossils or the rock cycle, even though these topics are commonly taught to students at this age level.

By eleventh grade, students in Pennsylvania should have had two exposures to earth science, one in grade 6 and another in grade 9, although none of these students mentioned having learned about earth science in grade 6. These students would have reviewed information taught in grade 6 and been taught principles of relative and absolute dating of earth materials. Students at this level will have also been exposed to instruction regarding the Big Bang and the nebular hypothesis.

Place value, proportional reasoning, and scientific notation are introduced to students in Pennsylvania prior to 8th grade. All students in this sample would have had some experience with these topics prior to the interview.

3.6 Methods of data analysis

Since this is an exploratory, qualitative study, data analysis was primarily qualitative. The small sample size coupled with the way participants were chosen lends itself to qualitative rather than quantitative data analysis methods. All interviews were fully transcribed. Transcripts were analyzed by the investigator to find recurring themes or common responses across participants. For example, did participants describe duration strategies that were consistent with the use of spatial mapping to temporal tasks? Did they demonstrate confusion about the size of numbers of various magnitudes? What types of reasons did they give for their answers? Did they rely on surface features to make judgments rather than geoscience principles? The goal of the analysis was to find patterns in responses that can be investigated in future studies. Students were sorted into groups based upon their responses for some items.

The investigator sorted numeric timelines into groups by accuracy on a linear scale using visual inspection of the timeline alone in accordance with a list of criteria. A second rater independently sorted responses according to the same criteria. Discrepancies between ratings were discussed, and all differences in categorisation were reconciled. Inter-rater reliability was established and will be discussed in chapter four. The investigator then compared transcripts with the timelines. In any case where the transcript suggested a more accurate conception than the timeline alone, the timeline was moved to a different category.

However, there was some quantitative data analysis for which there is precedent in qualitative research (Denzin & Lincoln, 2005). Quantitative data analysis was done for two reasons. First, several tasks in the interview protocol were adapted from items used by other researchers who applied quantitative methods to their data. The items in this study were analyzed in similar ways to allow for comparison with the data of others. If findings from this study corroborated those in previous research it could lend support to the conclusions of others. If current findings disagreed with those of earlier researchers, it may mean that new interpretations are required.

Second, the duration animations were designed to not only provide qualitative analysis of *why* students chose the answers they did but also to be able to investigate whether the order of presentation of the animations affected student performance. Thus, t-tests were calculated in this one instance to ascertain if there was a difference between groups depending upon which animation was viewed first. Effect sizes were also computed. This small amount of statistical analysis was used in tandem with the qualitative analysis to reach conclusions about the data.

CHAPTER FOUR

RESULTS OF THE STUDY

This chapter is organized in the same manner as the literature review and corresponds to the three “legs” of the “stool” described in chapter two. All tasks in the interview protocol dealt with either succession or duration. Each was described in chapter three. Some tasks were designed to investigate one of the three “legs” alone: conventional time, large numbers, or geoscience content knowledge. Others were created to probe two “legs” simultaneously. (Figure 3.1 on p. 170 lists interview items and “legs” to which each referred.) Results for each “leg” of the “stool” are discussed separately and then compared with one another.

4.1 Conventional time and a concept of deep time

According to the model of the “three-legged stool,” an understanding of deep time depends to some extent upon conventional time conceptions. What would happen if we created succession and duration tasks that were set in a geologic context but allowed them to occur in conventional time? As stated in chapter three, these tasks were not created to model any geologic process per se, but rather to assess understanding of conventional time using tasks that could be adapted to ask a similar temporal question in a geologic context. If difficulties conceptualising deep time can be partially accounted for by a poor conception of conventional time, we would expect to see certain things. Overall, we would anticipate few difficulties with conventional time concepts since the students in this sample are all adolescents or adults. However, there may well be *some* students whose understanding is marginal.

Students who have trouble with succession may be unable to successfully sequence a series of events in real time. They may perceive simultaneous events as successive. Those who have difficulty with duration are likely to make some of the same errors described in chapter two. They may equate size with duration, e.g., greater size equals longer duration. When duration can be judged on the basis of size or rate alone, they may successfully complete the task. When they must take account of both factors at the same time, they will be less able to judge accurately. Results for succession and duration are discussed in turn.

4.1.1 Tasks exploring understanding of succession

The succession tasks in this study were of two types. The first involved the before and after relationship in which one must determine which of two events preceded the other in time. Students were asked to determine which of two fossil layers formed first after watching an animation of those layers appearing in order. A still image of the animation is in Figure 3.2 (p. 172). The second is an extension of that idea. Participants were asked to sequence all of the fossil layers correctly. This involved not only before and after relationships but also simultaneity since the same fossil layer could be found in multiple places in the columns and appeared together. If students do not possess a clear understanding of succession they could be expected to be unable to determine which of two layers appeared first and/or be unable to accurately sequence the layers.

4.1.1.1 The before and after relationship

After viewing the animation as many times as they wished, students were asked three questions in which they had to compare the relative age of two fossil layers. Participants had little difficulty with these questions involving the before and after relationship. Ninety-seven percent of the students (34 of 35) answered all three questions correctly. An eleventh grade girl initially answered one of the three questions incorrectly, yet when sequencing all the fossils she ordered those same layers correctly. When the discrepancy was pointed out to her, she self-corrected and was counted in the 34 who answered all questions correctly. An eighth grade boy was the only one who struggled with this task. He answered only one question correctly. He continued to have trouble with the next task, and his responses are discussed more fully in that context in the following section.

4.1.1.2 Simultaneity along with the before and after relationship

Students were next asked to place all fossils in order, a task modelled after Puzzle 5 in the GeoTAT used by Dodick & Orion, (2003a, 2003b) which was described in section 2.6.1.1 and can be found in Figure 2.3 (p. 132). The difference is that their task only used the static sequence shown to the students in the present study prior to watching the animation (discussed in section 4.3.1), while in this study students actually saw the fossil layers appear on the screen. Eighty-three percent (29 of 35) of the students placed all fossils accurately, slightly less than the 34 out of 35 who answered all three comparison questions correctly. Thus, sequencing all the layers proved to be a more difficult task. Three of the 29 originally placed some out of order but later self-corrected. Six students (four eighth graders, one eleventh grader, and one university student) did not order the fossils correctly. Most switched one pair of fossil layers. Jamal, the student who only answered one question correctly in the previous task placed only 20% of the fossils correctly.

In order to compare these results with those of Dodick and Orion (2003a, 2003b), similar data analysis methods were employed. A full comparison of their results and mine can be found in chapter five. There are a total of ten different fossils in the exposures. Participants were scored based upon the number of fossils ordered correctly. Number correct was transferred to a percentage. Mean number correct for each grade level was calculated. Means and standard deviations are reported by grade level in Table 4.1. Mean number correct at each age level is quite high, and standard deviations are relatively low. The exception is the relatively larger standard deviation for the eighth graders. This is attributable to Jamal's very low score which is described below. As already indicated, the sample for this study is quite small and not randomly selected. Thus, there is no assumption that these results would be representative of the population of 8th and 11th graders and university students.

University		Grade 11		Grade 8	
Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
99	2.8	98	5.8	89	20.2

Table 4.1 Comparison of mean grade scores and variance for succession task (N =35)

In explanations for their ordering students described the before and after relationships and simultaneity among the layers, i.e., the principle of fossil succession without giving any evidence they were aware they were applying a geologic principle. At least nothing in what they said prior to or after the task would lead one to conclude they were familiar with the term fossil succession. Many noted that the columns filled in order from the one on the left to the one on the right in what they described as a diagonal pattern or a pyramid. Ayanna's explanation exemplifies the thinking.

> *I knew that anytime two repeated even if they were in different columns they appeared simultaneously so the ammonite went on top on the left but it's also in the middle so I did that. Then I did the coral and then clam cause I remember that as the first one was filling up the 2nd one started the clam. Then you finish that column with the gastropod which also repeats in the middle. From there you go up one to the brachiopod but that's on the bottom of the 3rd so now not only do you have the 4th in the middle but you have the 1st one on the right. Then it didn't skip so I had to go to fish scale, snail cause it wasn't like just one floating in the air on top of that. Then you finish it up with the sea urchin and the shark tooth.* (Ayanna, 11)

Only one student actually used before and after terminology in his explanation.

> Interviewer: Tell me what you're doing when you're giving me those in order.
>
> *Sean: I'm looking at what comes before and after it cause you can almost put these into a certain order by looking. You can use each row [column] to find out what the next one is by saying, ok the trilobite, then there's an ammonite [1st column]. There's also an ammonite there [2nd column] so there can't be anything before that. Then you say, ok there's a clam in the second and in the first but before the clam is a coral although it's not in the second row [between] ammonite and the clam, so I know that there's a coral, then I know*

that there's nothing else before the clams. Then there's nothing else before the gastropod, and then there's the brachiopod cause obviously the snail can't come before the brachiopod by looking at that row [2nd column] and this row [3rd column] so then before the snail, though, there is the fish scale. Oh, I think I messed up on that one, now that I think about it, but it was the fish scale and then there was the snail.

Interviewer: That's the order you put them in.

Sean: Oh, I did, ok. Then after the snail I could tell that the sea urchin and the shark tooth cause there's nothing else, there's no other row to look at. (Sean, 11)

Four learners made an error similar to one found by Piaget (1969) with much younger children in that they said two fossils that appeared simultaneously actually appeared one after the other. Piaget reports on a task in which two lamps were set up some distance apart from each other and the child, with different arrangements employed in various trials. The child had to determine which of two lamps was lit first with a one-to-two second difference in when they were lit or if both were lit at the same time. Younger children often answered incorrectly and said the two were lit successively when it was simultaneously, or vice versa. By the age of 10-11-years, those errors were very uncommon. However, in the present study a similar error appeared. Despite the fact that two of the four students who made this error correctly ordered the fossil layers, all indicated that similar fossils did not appear simultaneously, but rather one after the other. Alyssa (8) exemplifies the point.

Interviewer: I'd like you to tell me the fossils in order from the one that formed first to the one that formed most recently.

Alyssa: Trilobite

Interviewer: What came next?

Alyssa: Ammonite, on the second column, though because it was this one [2nd column] and then this one [1st column]. And then ammonite same one.

Interviewer: Ok, were they at the same time or did this one [2nd column] come before this one [1st column]?

Alyssa: This one [2nd column] came before this one [1st column].

Interviewer: Ok

Alyssa: Then it would go ammonite again but on top of it [trilobite].

She went on to say that the clams and the gastropods appeared one after the other rather than simultaneously. Oddly, however, she indicated that the brachiopods appeared at the same time. Two of the students in this group, (Sofia, 8 and Nathan, 11), initially said that the two ammonites appeared successively but after further questioning expressed confusion and finally decided that they must have appeared simultaneously. Participants' ocular movements were not tracked to determine which of the two pictures they judged to appear first. Therefore, it is not possible to say if there is a pattern in which of the two pictures they said appeared first. It can be said that at least some older students have difficulty judging the simultaneity of events. A learner's visual perceptual abilities may impact performance on this item.

The only person who had problems with this task was Jamal (8) who answered only one of the three questions correctly after watching the animation. When sequencing the fossils he started with the gastropod. When asked about the appearance of the gastropod relative to the ammonite he confirmed that he saw the gastropod appear before the ammonite that was below it in the first column [which was not correct since each column filled from bottom to top]. Jamal seems to have serious difficulty dealing with succession in conventional time. It is not clear if that is an artefact of the experimental design or if he truly has a conception of succession that is typical of a much younger child.

4.1.1.3 What do these students understand about succession?

What can we conclude about how this group of students understands succession in conventional time? With a few exceptions, this sample appears to have a good grasp of the concept. They had little or no trouble judging the before and after relationship or simultaneity. The geologic context itself proved to be no impediment to their responses. Answers were largely consistent across tasks. That is, students who got all three questions correct also correctly ordered the fossils. There were a few exceptions. Five students (three 8th graders, one eleventh grader, and one university student) answered all three questions correctly but mixed up the order of several fossils when sequencing them. Inconsistent responses are often difficult to categorise. It could be that errors represent a lack of attention to the task. Conversely, these students may be at the point in their understanding at which the concept is not solid and thus, they sometimes answer correctly and sometimes incorrectly. Only one eighth grader appeared to have a poor concept of succession.

His explanations were typical of the younger children in Piaget's studies who responded in ways that contradicted what they had just seen.

Succession is only one component of a concept of time. A person must also be able to judge the duration of an event using various cues.

4.1.2 Tasks exploring understanding of duration

As was demonstrated in chapter two, duration of two events can be judged in several ways. If starting and ending times are available, they can be used. When that information is not available (and we are talking about motion), a person must weigh rate against distance to judge. In this portion of the interview, students watched three animations designed to investigate their understanding of duration (p. 175-176). The animations were different from one another in the following ways:

- Animation 1 (A1): layers all same size; different rates, different durations
- Animation 2 (A2): layers different sizes; different rates, same durations
- Animation 3 (A3): layers different sizes; different rates, possibly different durations

Presentation of animations was counterbalanced so half the students saw A1 first and half saw A2 first. A3 was always shown last. After each animation students were asked two questions. The first question asked which layer took longer to fill while the second question asked which layer filled faster. (See Table 3.3, p. 178 and Appendix A for correct answers.) As with the succession animation, these animations were viewed on a laptop set at a distance of each participant's choice. They were able to watch each animation multiple times both before and after the questions were asked.

4.1.2.1 Results of duration tasks

Table 4.2 presents total number of correct responses across all three animations and allows for comparison of responses across age levels. The table gives a broad overview of student performance but does not indicate how many correct responses a student had for each animation. For example, someone could get three correct by answering one question correctly from each animation or by giving two correct answers for the same animation and another correct answer for either of the two remaining animations. In contrast to the succession tasks in which students performed quite well, they did not perform as well on these tasks as some of the research, particularly Piaget's (1969), described in chapter two would have predicted. There is only a marginal difference in mean number of correct answers between 11th graders and university students and none between 8th and 11th graders. Additionally, the only two students to answer all six questions correctly were eighth graders. If there is any age effect at all in this data, it is very small.

	1 correct	2 correct	3 correct	4 correct	5 correct	6 correct	Mean number correct
8th grade	0	4	4	0	2	2	3.5
11th grade	0	1	4	6	0	0	3.5
University	1	1	1	7	2	0	3.7
Total	1	6	9	13	4	2	3.5

Table 4.2 Total number of correct answers to animation questions, all participants (N=35)

Another way to look at the data is to compare responses for each animation. This comparison can be found in Table 4.3. This table indicates how many questions students answered correctly after each animation depending upon which animation they saw first. Half (N=18) saw Animation 1 first and the remaining 17 saw Animation 2 first. This data is not broken down by grade level since age appears to be a small or nonexistent factor. While a majority of students got at least one question correct for Animations 1 and 3, they did not do as well with A2. Since size and rate both varied, it was expected that students would find A3 to be the most difficult, but this was not the case. At all age levels students did better on this animation than they had on the first two. There are several possible explanations for this, one of which is a practice effect which will be discussed more fully in section 4.1.2.4. The number of correct responses for A3 was similar no matter whether a student saw A1 or A2 first. That was not true for Animations 1 and 2, although the trend is not quite as strong for A1 as it is for A2.

Students were randomly assigned to the group that saw A1 or A2 first. There is no reason to assume any systematic difference between the groups. Therefore, a t-test for independent means was calculated for each animation to see if the difference in number correct based upon which animation was seen first was statistically significant. For A1, $t(33) = 1.74$, $p < 0.10$, while for A2, $t(33) = 2.20$, $p < 0.05$. Effect sizes were also calculated for A1 and A2 since they are not dependent upon sample size. For A1, $d = 0.43$, which is a medium effect size but significant only at the 0.1 level. A large effect size ($d = 0.86$) which was significant at the 0.05 level was found for A2. These results indicate that watching A1 first helped interviewees with their performance on A2 but that watching A2 first seemed to have a modest negative impact on their performance on A1. There was no statistical difference between groups on A3 depending upon which animation was watched first, [$t(33) = 0.91$, *ns*]. A small, non-representative sample requires caution in the interpretation of these results. They do, however, suggest further ways in which duration might be explored, a point that will be discussed in chapter six.

As a group, students found it much easier to deal with a situation in which the size of the layers was the same but their durations were different (A1). In this case, there was a direct proportional relationship between rate and duration. A student only needed to consider the difference in rates to correctly determine duration. Size of the layers was immaterial since it was held constant. They found it more difficult to deal with questions in which the size of the layers being compared was different while their durations were the same (A2). In that case there was an inverse relationship between the size of the layer and the rate at which it filled. Those two pieces of information had to be reconciled with one another in order to conclude that durations were equal.

Animation	Which one seen first	0 correct	1 correct	2 correct
Animation 1 (A1)	Saw A1 first (N=18)	1	5	12
	Saw A2 first (N=17)	3	7	7
	Total	4	12	19
Animation 2 (A2)	Saw A1 first (N=18)	8	6	4
	Saw A2 first (N=17)	14	2	1
	Total	22	8	5
Animation 3 (A3)	Saw A1 first (N=18)	2	5	11
	Saw A2 first (N=17)	1	3	13
	Total	3	8	24

Table 4.3 Correct responses to animation questions by which animation was seen first

4.1.2.2 Explanations given for responses

Students cited a variety of reasons for their responses, and many mentioned multiple reasons across their explanations for the six questions. All the explanations reported in this section come from A2. This is purely coincidental and means only that the explanations chosen as clear, illustrative examples of particular strategies happen to come from A2. Perhaps this is because A2 required a coordination of rate and size to judge duration while A1 did not. Therefore, the explanations were more indicative of the thinking necessary to judge durations when there was competing information. Explanations for the other animations were not deliberately excluded.

The size of the layers, the perceived speed at which the layers were filling, or the use of counting (either in their heads or by using the timer on the animation) were named as strategies with basically equal frequencies when considering the animations as a group. If a student focused exclusively on the size of the layers, judging durations for A2 was problematic. This type of strategy would have been successful if rates had remained constant. If rates are constant then the size of the layer would indicate which takes longer to fill. Since rates were different, equating size alone with duration resulted in an incorrect response. This is, indeed, a type of error that was predicted if students have difficulty judging duration in conventional time.

Alyssa (8) saw A1 first and explained her answers by saying one layer moved faster than the other, using a strategy that was sufficient for A1. Yet, on A2 she switched to a size-only explanation.

> *Cause the blue layer has more levels and stuff to it than the brown layer.* (Alyssa, 8)

Some students like Ashley (8) seemed surprised by the question as she emphatically asserted, "*The pink one cause it's shorter.*"

Size-only explanations were used by older students, not just 8th graders.

> *Cause there is more and I guess that since it's more it would take more time to fill up then.* (Malik, 11)

To use perceived speed to obtain an answer, an individual had to weigh the perception of rate against the relative sizes of the two layers in question. Some students were aware that they needed to take account of both size and speed but were unsure how to do so. Not all were able to successfully reconcile the two pieces of information. Evan's (8) explanation for a question in A2 exemplifies the dilemma.

Evan: The yellow did take longer. It just seemed to be moving a little slower compared to the blue. The blue appeared to be just speeding along.

Interviewer: The blue's a lot bigger than the yellow.

Evan: I know, I'm thinking, I'm thinking maybe that has some sort of connection. [He was unable to elaborate.]

Justin (11) tried to account for both pieces of information but was similarly unable to connect them properly.

Interviewer: Did the blue layer or the yellow layer take longer to fill up? [Yellow is correct.]

Justin : The blue layer took a little bit longer and just from watching, like, if the yellow was probably proportional with the blue size then the yellow would have taken longer but I think the blue just took longer to fill up.

Interviewer: Let me see if I understand what you're saying about the blue and yellow being proportional. Tell me a little more about that.

Justin: The blue was moving faster than yellow was but the yellow had less space to go so [pause]

Interviewer: That's why the blue took longer?

Justin: Yeah.

Another 11th grader, Nathan was not merely conflicted about how to reconcile two competing pieces of information for A2. He had a third piece—he had counted and knew that both layers took six seconds to fill. Yet, the fact that the blue layer was filling faster (It was also larger), led him to conclude that the yellow must have taken longer to fill. He qualified his answer by describing it as "slightly" longer.

Interviewer: Did the blue layer or the yellow layer take longer to fill?

Nathan: I think they were both around 6 seconds.

Interviewer: So you think they were about the same or did one take longer?

Nathan: I think yellow might have been just slightly longer.

Interviewer: Do you want to watch it again?

Nathan: Yeah.

Interviewer: If you think they're the same you can say that as well.

Nathan: The question was which one takes longer? Ok, the blue just moved a lot faster. The yellow was a lot slower, sort of around the same time but I think yellow was just longer.

Interviewer: You're saying it seemed like it was around the same time because you were counting?

Nathan: Uh huh

Interviewer: But it also seemed to you like the yellow was slower?

Nathan: Yeah

Interviewer: So you think it must have taken longer.

Nathan: Well, it's not necessarily but since they're different sizes but [sic] I think it was just slightly longer.

It is fair to say from the excerpt with Nathan that he displayed some confusion about how to deal with competing pieces of information. It is possible that the interviewer bears some responsibility for his confusion. The first follow-up question, "So you think they were about the same or did one take longer?" could have led Nathan to conclude that his initial answer was incorrect. If he had been the only student who appeared bewildered by how to reconcile the different aspects of what he saw it would be inappropriate to make inferences based upon his responses. However, he was not the only one.

Sometimes the attempt to reconcile size with rate resulted in faulty reasoning that did not fit the animation and evinced thinking that was similar to those who simply said larger = longer duration. Elizabeth's (univ) explanation for A2 is novel. She appears confused as to why a smaller layer would take longer to fill, but attempts to solve her dilemma with an incorrect analogy.

Interviewer: Did the blue layer or the yellow layer take longer to fill?

Elizabeth: I think it was the yellow. Yellow was taking a long time.

Interviewer: Why do you say that?

Elizabeth: I just noticed that, but I don't know why it should take longer cause it's like half the size of the blue one. But I guess if you think about it, like, filling *up something that's empty, the yellow may be more deep or something so it maybe is like taking longer.*

Interviewer: Let me see if I understand what you're saying. At first you said it would seem like the yellow would go faster because it's smaller.

Elizabeth: Yeah, well, and like at the end of it you think it should have been but I noticed that it's slower.

Interviewer: What was the last thing you said about it being deeper? I didn't understand.

Elizabeth: I think if you think about taking a glass, or let's say like a bowl. One is more deep than the other, and one is wider and one is deep. You think that that [sic] one that's wider holds more, but it's actually shallow so it probably fills up faster. It may be like the yellow if it was a bowl or something, deeper one takes longer to fill. That's what I was thinking.

However, Elizabeth's analogy does not conform to what she saw in A2. She described the yellow layer as the deep one while the blue equated to the wider, shallower bowl. The latter might appear to have greater capacity but actually does not and therefore fills faster. Yet, in A2 both layers were the same width and the yellow one was shallower than the blue. Elizabeth appears to be indicating that capacity and duration must vary in the same way. In other words, size and duration are directly proportional. The visual perceptual information in the animation did not conform to her idea, so she reinterpreted that information to make it fit.

Durations could have also been judged by ignoring rate and size entirely and either watching the timer on the screen or counting to oneself. As a group, this strategy was mentioned more frequently for the second or third animation watched than the first one.

> *The blue layer, yeah the blue layer cause the blue layer takes 7 seconds and the red layer takes 5.* (Matt, 8)

Counting would appear to be an efficient strategy as it does not require someone to work back and forth from rate to size, but merely to watch the timer and engage in simple subtraction. Ironically, however, not everyone who mentioned a counting strategy correctly answered the items for which they said they counted. Ryan (11) mentioned counting as a strategy for Animations 2 and 3, yet both of his answers for A2 were incorrect and only one was correct for A3. Anthony (univ) said he timed Animations 2 and 3. He got both answers for A2 incorrect, but both answers for A3 were correct. In addition to Anthony, two other university students got answers incorrect when employing a counting strategy. The fact that a number of students answered incorrectly when counting could be interpreted in several ways. It could mean that some students have difficulty reconciling competing perceptual information (the number they got when counting and the size of the layers, for example). Nathan's response reproduced above lends credence to that interpretation. It could also indicate inattention to the task. The fact that some students chose to only watch each animation once despite multiple opportunities to see them again lends support to that inference.

4.1.2.3 Categorisation of responses by strategy employed

In order to compare cited strategies, participants were broken into two groups: those who answered three or fewer total questions correct and those who answered four or more correct. Dividing participants into more than two groups could not be justified based upon the small number of total responses (six) and the sample size. More groups would make distinctions between participants that are not justifiable since it was possible to understand the task and simply make a careless error. Sixteen participants answered a total of three or fewer questions correctly on all three animations. Six of those sixteen viewed A1 first while the remaining ten watched A2 first. Nineteen students answered four or more questions correctly on all the animations. Eleven of those individuals watched A1 first, and seven watched A2 first. Numbers reported in Table 4.4 list the number of students who cited each reason for their responses.

Reasons cited for answer	3 or fewer correct (N = 16)	4 or more correct (N = 19)
Size of layers	14	13
Pattern (alternating speed of adjacent layers)	3	4
Perception of rate (some said *seemed* longer)	16	15
Use of clock on screen or counting in head	8	16
Guessed or unsure why answer was chosen	4	1

Table 4.4 Reasons cited for answers to animation questions

Since students commonly cited multiple reasons across the six questions, column totals do not equal the number of students in that column. Overall, there is not a lot of difference between the two groups in terms of the reasons cited for their answers with two exceptions. The first is the number of students citing use of the timer or counting as a strategy. Almost all (16 of 19) of the students who answered four or more questions correctly mentioned counting at least once, while only half (8 of 16) of those who got three or fewer questions correct did so. Two of the students in the latter group said they were counting in their heads or using the timer on the screen, but with only minimal success. Seven of the eleven students who never mentioned counting as a strategy were eighth graders. Ten of those eleven answered three or fewer total questions correct across animations. However, the additional eighth grader who didn't mention counting as a strategy was one of only two students to answer all six questions correctly. Three of the remaining four students who never reported a counting strategy were university students. All of them got four questions correct.

The second difference between the two groups is not readily apparent from Table 4.4. Students in both groups mentioned size of the layers in their explanations but with an interesting difference. Those who answered three or fewer correct said the size of the layer determined how long it took to fill, i.e., thicker layers take longer. In contrast, when students who answered four or more correct mentioned size it was in the context of trying to weigh the size of the layer against the perceived rate at which it was filling. For them, size was a factor, but not the sole determiner of their answer as was typical for students in the other group. They were able to deal with the inverse relationship between rate and duration, while the former group was not.

4.1.2.4 Development of strategies

Many seemed to be trying multiple strategies as the tasks progressed and changing strategies with each animation. As mentioned, students were more likely to name counting as a strategy for A3 than for the other two animations. Students who employed a counting strategy for the first animation they saw tended to continue to use that strategy throughout, but not always with success. There appeared to be a practice effect as the group as a whole did better on A3 than the other two which was not expected. The first question following A3 was similar to A1 in that the two layers being compared (green and brown) were the same thickness. However, since they filled at different rates the amount of time necessary for them to fill (duration) was different. The second question was the only one of its kind asked during the interview. The two layers being compared (red and blue) were different thicknesses, filled at different rates, and took different amounts of time to fill. This question made an impression on a number of the participants as they cited this example later when trying to apply the animations to a stratigraphic sequence. The blue, thicker layer, filled much more quickly than the red, thinner layer. The perceptual difference in the rates was so apparent that it was difficult to ignore. More students answered this question correctly than any of the other five in this portion of the interview.

Hannah (11) is a good example of a progression of strategies from one animation to the next. She watched A1 first. Her explanation for the first question after A1 reminds one of travelling to some desired event and how it always seems to take longer to get there than it does to come home.

> *When I answered I was more thinking like the first time I watched it I didn't really know what was gonna happen and so I was waiting to see what would happen, so, of course, you know, that's gonna take longer, like it's gonna seem like it's a longer amount of time, but then like after I knew what was gonna happen, and like I watched it a second time it seemed like it took longer for the blue one to fill because I wasn't like, no, why is this taking so long, what's gonna happen? You know what I mean, like, once you know something's gonna happen you can see it the way it really is as opposed to what you are kinda like anticipating.*
>
> Next, she watched A2. Note how her explanation changes.
>
> Interviewer: Which layer took longer to fill—the blue layer or the yellow layer?
>
> *Hannah: The blue.*
>
> Interviewer: Why?
>
> *Hannah: Well, I started to pay attention to like the time, like the red one took 5 seconds, the blue was like about 7, and then yellow I kind of, it was kinda hard to watch at that point cause I was trying to go back and forth like from the time and actually watching the different colours changing. I think yellow was about the same as the red or like at least around there.*

By the time she got to A3, she appeared to be having no difficulty employing a counting strategy. When asked if the green or brown layer took longer to fill, she replied,

> *The brown one took about 10 seconds or approximately as opposed to green which took about 5 or 6.*

4.1.2.5 What do these students understand about duration?

Overall, there was very little difference in how students at different ages performed on this task. One participant, a female university student, got one question correct, while only two students (both eighth graders) answered all six questions correctly. The one age difference is that eighth graders were somewhat less likely to describe a counting strategy than older students. Everyone, with the exception of two eighth graders, reported using at least two different strategies. Fifteen students cited three different strategies as reasons for their responses, and four said they used four different strategies.

What can be said about how this sample understands duration in conventional time? That is a difficult question to answer. It is hard to allege that incorrect responses provide clear evidence that students *couldn't* complete the tasks successfully. Some seemed mixed up when explaining their answers but when offered the opportunity to watch the animation again they said they didn't need to and then simply gave an answer. Perhaps their performance says more about their motivation for the task than it does their understanding of duration in conventional time. In some cases there were specific hints that inattention rather than a faulty conception of time was responsible for errors. Consider Vincent (univ),

> *I guess I could have sat here and marked, oh, that one took 2 seconds and this one took 4 or 5 and been a little bit more observant.*

or Ashley (8) after A3,

> *I really paid attention this time.*

Some students chose to watch each animation only once even after being given opportunity to see it again. While this may have affected the accuracy of responses, the number of times a student watched the animation was not a good predictor of a correct answer as students who watched the animations several times also got answers incorrect. The fact that participants reported changing strategies multiple times throughout the task makes it difficult to determine if any one strategy was more successful than another. This is further complicated by the fact that the use of a particular strategy did not consistently produce either correct or incorrect responses.

Many students spontaneously said two layers took the same amount of time to fill on A2. Even though that was not expressly stated as an option, they viewed it as a viable choice. Claire was one of two students who specifically asked whether or not that was a possibility.

> *Claire: The brown one, there was a difference in the rate at which it was filling up. I definitely saw the pink one was slower and this one [brown] was faster.*

The only thing I have to decide now is whether the difference in rate was enough. Is it an option to say they took around the same time?

Interviewer: Yes, that's an option.

Claire: I think they would have filled up around the same time. The brown one is pretty much double the size of the first one even though it was filling up faster it would be around the same cause the other one is a slower rate.

[answer is correct]

It could be that some of the others who answered incorrectly on A2 did so, not because they didn't know the correct answer but because they thought the correct answer was not an option.

Despite those caveats, some tentative findings emerge. These students had more problems judging duration when they needed to simultaneously account for rate and size than when duration could be judged on the basis of rate alone. When perceptual evidence disagreed with a current idea, the data was sometimes ignored as in the case of Ashley who seemed surprised that anyone would ask if a thicker layer could possibly fill more rapidly than a smaller layer. For others, like Elizabeth, the task was reinterpreted to fit an existing idea and what she described did not correspond to what had actually taken place in the animation.

4.1.2.6 Application of understanding of duration to a stratigraphic sequence

Ault (1980, 1982) reported that even when students displayed a solid understanding of succession in conventional time they were unable to apply that knowledge to a geologic outcrop. The next item was designed to see if a similar phenomenon would be observed with duration. After watching the animations and answering the six questions students were shown a drawing of the stratigraphic sequence in Figure 3.6 (p. 179) that is typical of what one might find in the Southwest United States. Students were asked to compare layers 3 and 4 (from the bottom) and say which one probably took longer to form based upon the animations they had just seen. There were four possible responses:

- layer 3 (thicker) took longer
- layer 4 (thinner) took longer
- both took the same amount of time
- can't be determined from the picture alone

The last choice is the correct answer since the picture alone gives no clue as to the depositional environment. Frequency of student responses can be found in Table 4.5. The animations were not specifically constructed as a teaching intervention as that was not the purpose of this study. Therefore, the interviewer did not engage participants in a discussion about general principles they observed from the animations that might apply to this task. Students were merely instructed to use information from the animations to help them answer the question about the stratigraphic sequence similar to Ault (1980).

Response	Total Frequency	Frequency for students who answered 3 or fewer correct on animations (N=16)	Frequency for students who answered 4 or more correct on animations (N=19)
Layer 3 (thicker) took longer	8	6	2
Layer 4 (thinner) took longer	18	10	8
Both took the same amount of time	3	0	3
Can't be determined from the picture alone	6*	1	5

* includes one student who listed all three of the other possibilities as answers.

Table 4.5 Student responses comparing time for two adjacent sedimentary layers to form

Despite the fact that students were instructed that the design of the layers and the erosion patterns had nothing to do with how long the layers took to form, some, like Ben (8), initially focused on those features in their answers.

Ben: I think the top one [layer 4].

Interviewer: Why do you think that?

Ben: I guess cause it's smaller and like there's a big chunk out of it right there so I think it woulda filled up quicker cause there's a big chunk missing.

Students who answered three or fewer correct were more likely to equate greater size with greater duration required for deposition. Those who said layer 3 took longer all focused on the relative size of the two layers in their reasoning. Of the eight students who said 3 took longer, five of them had invoked size as a reason for one or more answers following the animations. Two of the others justified their responses to the animation questions by saying they counted. Some like James (11) who said larger took longer said it was consistent with what they had seen in the animations, although that was not always the case,

Because the bigger ones usually took longer than the smaller ones

A few like Ryan (11) who said the larger layer took longer indicated their answer was not based on what they saw in the animations at all.

Ryan: I'd say this one [layer 3] took longer to form.

Interviewer: The thicker one?

Ryan: Yeah.

Interviewer: And why would you say that?

Ryan: Just cause of the size, I guess.

Interviewer: Is that based on what we saw in the movies or based on something else?

Ryan: That's based on just how long I would perceive it would take to make. I figure if the average person's gonna make a big ball as opposed to a small ball, it would take longer to make.

Interviewer: Was that true in the movies?

Ryan: Not all the time.

Students were most likely to say layer 4 (the smaller one) took longer to form. This was true regardless of how many questions an individual answered correctly after the animations. That response was unanticipated and unfortunate as one of the animation questions may have instilled an alternative conception about sedimentation in the minds of some participants. The students who said “smaller took longer” appeared to be relying exclusively on the comparison of the red and blue layers from A3 (described in 4.1.2.4) for their explanations. Even though the blue layer was almost three times as thick as the red layer, it filled two seconds faster than the red. The rate difference between these two layers was greater than for any of the other animation questions and appeared to make a significant impression on many of the participants. When probed, some acknowledged that “smaller took longer” was not true in all the animations. Others like Evan (8) were unsure but ultimately decided all three animations demonstrated “smaller took longer,”

> *Evan: I’m saying the smaller one, top one might take longer to form cause it just seemed like in the movie that the slower ones seemed in general to go slower than the faster ones, I mean the larger ones, like the last one the blue was larger but it was just speeding along and so I’m gonna say that one.*
>
> Interviewer: Was that true in all the movies that the shorter ones took longer than the bigger ones?
>
> *Evan: I’m trying to think, uh yeah I think so, I’m not sure, yeah.*

Still others, like Michael (11) were more confident that “smaller took longer” was indeed true for all three animations, although that was not correct.

Michael: The thinner one took longer.

Interviewer: Why do you say that?

Michael: Because in all the movies the thinner ones took longer to fill.

Three students said both layers probably took the same amount of time to form. Once again Claire (univ) appeared to be trying to coordinate information about a size difference with a rate difference as her response demonstrates,

Claire: Does that relate to any one video or just in general?

Interviewer: Just in general to the movies.

Claire: I think in general if you have a layer which is like half the size of the other layer, the bigger ones were generally faster, came up faster and the smaller ones were a lot slower, but I think it would have taken around the same time because of the size again. One is almost double the size and that would kind of compensate for the rate at which they accumulate.

Sean's (11) response illustrates a point that cannot be overlooked when interpreting interview data. Although he ultimately ends up concluding the two layers took the same amount of time, he appears to be attempting to figure out the intent of the question and base his answer upon what he thinks the investigator is looking for. He says that in the absence of the animations, his first conclusion would be thicker layer = longer duration. He is also aware that the animations didn't consistently bear that out. His first response was to say layer 3 (the larger one) took longer to form.

Interviewer: What about in relation to the movies?

Sean: Looking at that I would definitely question it more and think it would be closer to, like they would be closer to the same amount, but, I mean, then I see this one [layer 3], this other one of the same type [layer 1] and it kinda makes me think differently about it as well with this one [layer 3] and that one [layer 4] because that one's [3] bigger than that one [4], so I would think [pause]

Interviewer: Suppose the other layers weren't here. Just suppose it's those two [layers 3 & 4], then what would you think based upon the movies?

Sean: I would think they'd be about the same.

Interviewer: Why do you say that? How did the movies help you decide that?

Sean: I think just cause I felt more like, I don't think it's physically the fact that it is that, I think it's just the matter of the fact, that, you know, I'm thinking since I've had that question asked to me and I feel like, mmm, maybe that is what they're looking for. You know, this long is the same as this long but we're not realizing it. I don't know, I'm trying, like, yeah, I just think based upon what I saw in the movie and the way it sounded, like maybe I was right, maybe I was wrong, looking at the time and everything, yeah, they'd probably be about the same.

Interviewer: And if you hadn't seen the movies?

Sean: Oh, I would have definitely picked the bigger one.

Six students said the answer to the question can't be determined from the drawing alone. While five of those individuals had four or more correct answers on the animations, one of the six answered three animation questions correctly. Three of the six said the answer to which of the two layers in the sequence took longer to form "can't be determined" in addition to one of the other choices in the course of their explanation. The sixth person was placed in this group because he listed all three of the other choices as possibilities and did not choose among them. Peter (univ) is an example of someone who mentioned another choice but ultimately concluded it "can't be determined."

Peter: A safe assessment would be to say that the thicker one took longer to form because it took more time for the deposit, for the sediment to deposit, but then it also depends on how fast the water was moving over the area where it was getting deposited so it could take the smaller layer the same amount or even longer just depending on how fast the water above it was moving.

Interviewer: If I understand you correctly, while you might think it would be this one [layer 3], it could actually be either one of them.

Peter: Right and there's no way from this to tell.

David (univ), the lone university geology major in the group, indicated his response was based upon the animations, and seemed to feel that anyone could reach the same conclusions he did.

David: Can't really tell.

Interviewer: Why do you say that?

David: If it's based on the animations some of the larger ones formed a lot quicker than the smaller ones and that could be the same thing in sedimentary rock formation.

Interviewer: Is that something you got out of the movie or did you know that before?

David: That's something I got out of the movie, but it's not like it's a hard concept to grasp. I mean, just because it's larger doesn't mean it took longer to form.

What are we to make of students' responses? As with the questions following the animations themselves, there may have been some who felt they had to choose one layer over the other, i.e., that "can't be determined" was not an option. There are hints of this phenomenon from several individuals as Justin (11) shows. He is an example of a student who offered two other responses before finally settling on "can't be determined."

Interviewer: Based upon what you saw in the movies what could you say about which of those layers probably took longer to form?

Justin: Well, I guess it could be different cause sometimes the small one would take longer than a larger one but then, if I was just, if I was a geologist, I would just assume the longer one took longer just cause it's bigger, but it could be different based on the videos. If I had to guess I would take the larger one.

Interviewer: You would guess the larger one based on the size. What if you just had to base it on the videos?

Justin: Based on the videos I'd probably say maybe the smaller one just cause it seemed like if it's smaller then sometimes it would take longer than like the larger ones would just go faster and then stop and then. On the basis of the video it seemed like if it was smaller it could go longer so, yeah.

Interviewer: Let's suppose I gave you a third choice—the first choice is "the bigger one took longer," the second choice is "the smaller one took longer," and the third choice is "you can't tell." What would you pick then?

Justin: I would say you can't tell by just looking at it.

One of the eighth graders (Connor) who answered all six questions correctly following the animations also correctly answered the stratigraphic sequence question. The other (Kayla) did not. She initially said the larger layer took longer. When asked if she was basing her answer upon the animations, she said she wasn't because the thinner layers took longer in the animations [not always true]. She then indicated that the smaller layer took longer to form.

4.1.3 What can we say about how conventional time impacts an understanding of deep time?

The first research question asked whether students make similar errors on tasks involving conventional and deep time and whether they apply similar strategies to both. Students in this sample showed little difficulty in performing succession tasks in real time. They were almost uniformly capable of determining the before and after relationship between two events and using simultaneity to correctly order a series of events. In this study the geologic context itself, i.e., the sequencing of fossil layers, did not appear to impede a student's ability to sequence events. One could argue that showing students the formation of fossil layers in real time removes the salient piece of geoscience content knowledge (principles of stratigraphy).

Performance on the duration items stands in contrast to the succession items. Here results raise as many questions as they answer. Many students did not perform uniformly on the items dealing with duration, and there is only the merest hint of an age effect. For whatever reason(s), participants did not consistently demonstrate a solid grasp of duration in conventional time. Possible explanations for this phenomenon will be explored in chapter five. Typically, individual students cited different strategies for different animations. It is unclear whether students changed strategies due to something inherent in the animations themselves or for some other reason. While some students never cited counting or the use of the timer as a strategy, a few mentioned it after every question. It was more common for students to mention counting later in this portion of the interview. Generally, a counting strategy was more likely to produce a correct answer, but that was not always the case as was shown in section 4.1.2.3.

Since the animations were not designed as teaching interventions, there was no discussion following them of any general principles students discovered or how those might apply to a geologic context. When comparing the two layers in the stratigraphic sequence, students were more likely to say the smaller layer took longer to form than the larger layer because that is what they observed in the animations. Yet, that was only true for one pair of layers they were asked to compare in A3. In the case of that pair, the difference in rate was so significant that it appeared to make an impression on many participants. Even students who correctly concluded that it was not possible to determine which layer took longer to form from the drawing alone often suggested a different answer in tandem with the correct one.

4.2 Large numbers and a concept of deep time

Another potential stumbling block to understanding deep time is an understanding of large numbers. Participants constructed a series of timelines for this study, a strategy employed by others (e.g., Confrey, 1991; Dehaene, 2003; Petitto, 1990; Siegler & Opfer, 2003) and described in chapter two. The main difference here is that students were asked to indicate the durations of various time periods in relation to each other as opposed to placing Arabic numerals on a number line. If a poor conception of deep time can be at least partially accounted for by a poor understanding of large numbers we would expect to see several things. First, since everyone in the sample is an adolescent or adult, most students should be able to successfully construct a linear timeline using familiar numbers up to 100. When asked to construct a similar timeline with large numbers we would expect that fewer students could successfully complete the task. We would anticipate greater logarithmic mapping with large numbers. Additionally, some students are likely to be confused about the magnitude of the numbers themselves and the proportional relationships between them. A timeline that requires students to place numbers that range from common human timescales to those in deep time should be especially difficult. Larger units of time will be less meaningful than smaller ones. Students can be expected to reason with units of days and months but find it harder to do so with thousand and million year units.

A description of the timelines appears in chapter three in Table 3.4 (p. 181). They are briefly described here for the reader's convenience. The first two timelines each contained four numbers. They were similar in several ways. First, both involved only four numbers. Timeline 1 (TL1) contained three time periods that everyone has experienced and one that no one in the sample has but that they would have all heard of. All four numbers in Timeline 2 (TL2) were outside the ability of a single individual to directly experience. Because the numbers in TL2 were all powers of ten, the timeline could show how students dealt with those proportional relationships.

Timeline 4 (TL4) had seven numbers that ranged from small (one minute) to very large (100 million years). TL4 was different from the other two in that it contained seven time periods that ranged from those in conventional time (one minute) to those in deep time (100 million years). The scientifically accepted durations of the events in Timeline 3 were rounded to the nearest power of ten or nearest unit of conventional time to allow for easier proportional reasoning and to allow for comparison of Timelines 3 and 4. A solid comprehension of deep time requires some understanding of the relationships among numbers that are very different from one another. It's useful to know that ten thousand years and ten million years are long periods of time. However, it is an incomplete picture. Some notion of the proportional relationship between those units provides a clearer image of the time required for various geologic processes. It also provides some notion of the relative amounts of time from the formation of the Earth to the appearance of first life versus the extinction of dinosaurs to the present. To save time students were only required to write the letter of each stimulus item for TL4 and not the time period itself on the timeline. Table 4.6 lists the items for TL4 with their accompanying letters.

Letter	Time period
A	10 million years
B	1 minute
C	1 year
D	1 month
E	1 day
F	10,000 years
G	100 million years

Table 4.6 Stimulus items for Timeline 4

4.2.1 How timelines were categorised

The three timelines were scored based upon whether or not students produced a linear timeline in which proportional relationships amongst numbers were clear. All timelines were sorted by the researcher prior to listening to the recorded interviews based upon similar criteria. Each timeline was scored as *correct, partially correct,* or *incorrect* based upon the criteria listed in Tables 4.7 and 4.8. Criteria for Timeline 4 are reported in a separate table due to the need to assess a linear, proportional scale for seven numbers that differed by many more orders of magnitude than was the case for the four numbers in Timelines 1 and 2. One table was insufficient to explain the categories adequately for four versus seven numbers. Criteria were chosen so as to allow initial sorting based solely upon the relative placement of all the events on the timeline on a linear scale. In the case of TL1, this would mean that 100 years would be at the extreme right side of the paper. One day, 1 month, and 1 year would be clustered very closely together at the left side of the paper. This demonstrates an understanding that in the context of 100 years, there is little difference between 1 day and 1 month. An attempt to create a timeline that shows a clear distinction between 1 day and 1 month serves to severely under-represent the difference between 1 year and 100 years. Latitude was given in scoring categories to account for perceptual-motor considerations. In order to be counted as correct, a student needed to create a timeline that placed all of the time periods on the same linear scale with reasonable accuracy.

Category	Criteria	Students in Category—Timeline 1	Students in Category—Timeline 2
Correct	Clear attempt to order times proportionally; first 3 numbers clustered at left side of timeline. Largest number is at or near right side of line. Third number is about 1/10 as far from left side of paper as 4th number is from 3rd number.	21*	16*
Partial	Some attempt to order times proportionally; distance between first 2 numbers is clearly smaller than distance between 2nd & 3rd numbers. Distance between 3rd and 4th numbers is too small. Third number is no more than 1/3 from left side of line.	6	1
Incorrect	No evidence or very minimal evidence that student was attempting to order times proportionally. One or more of the following is true: Numbers are spaced fairly evenly across the line. The third number is at or near the halfway point of the line. Proportional relationship among first 3 numbers is incorrect (i.e., distance between first 2 numbers is greater than distance between 2nd & 3rd numbers).	8	18

*Indicates one student whose responses were very difficult to classify. He is discussed in section 4.2.6.

Table 4.7 Sorting criteria for Timelines 1 & 2 with number in each category (N=35)

Category	Criteria	Students in Category
Correct	First 5 letters very close together at left side of timeline; G near right end of timeline; A no more than 1/4 of way from left edge	10*
Partial	First 4 letters very close together at left side of timeline; proportionality lost with letters F, A, & G	8
Incorrect	No evidence or minimal evidence that the student was trying to order times proportionally; generally numbers are spaced fairly evenly across the timeline	17

*Includes one student whose responses were very difficult to classify. He is discussed in section 4.2.6.

Table 4.8 Sorting criteria for Timeline 4 with number in each category (N=35)

A second person independently sorted the timelines using the same criteria, again without listening to the interviews. Since there were three categories into which timelines were sorted, two coders could be expected to agree 1/3 of the time by chance alone. Initial inter-coder agreement was 80%, 89%, and 91% for Timelines 1, 2, and 4, respectively. All three are well above what would be expected on the basis of chance ($\chi^2 = 12.6$, $p < 0.001$). Individual ratings were discussed and all differences in categorisation were reconciled by the two coders. The investigator then listened to each interview to determine if a student's verbal description of the timeline suggested a more sophisticated understanding than was indicated by the visual inspection of the timeline alone. In six cases for TL1, two for TL2, and one for TL4, a student's explanation suggested a clearer understanding than the timeline alone. In those cases, the timeline was moved to a higher category than it had been based solely upon visual inspection. An inaccurate timeline could be explained by a lack of interest in being precise, spatial reasoning difficulties, or fatigue on the part of the student. (After moving numbers several times some students said even though the final spot wasn't exact, it was good enough.) If this results in a systematic error it should be one that *underreports* a problem with large numbers. One purpose of this exploratory study is to determine whether student difficulty with deep time can be at least partially accounted for by a difficulty with large numbers. A systematic error that underreports the extent of the problem but still finds that a problem exists would lend credence to the notion that the phenomenon is real.

Many responses were easily categorised as correct or incorrect. A subset of the timelines did not fit neatly into either of those categories but seemed to represent either partial understanding of large numbers (or perhaps one of the alternative explanations mentioned in the previous paragraph). The chief difficulty with the partial category is that it includes a wide variety of responses from those of students who just barely escaped being classified as incorrect to those of students who were very close to producing a correct timeline.

In order to determine what the timelines might indicate about how these students understand large numbers, pupils were grouped by their scores on the timelines. There were seven different categories generated which can be seen in Table 4.9. The distinctions made among students by the use of seven categories are undoubtedly too fine for several reasons. First, they are based upon only one timeline of each type. The inability to draw a correct timeline might say more about a student's perceptual motor skills or attention to the task than it does about a concept of number. Secondly, the criteria used to define categories mean that some timelines are very close to the borderline of being placed in a different category. One might just miss being categorised as correct while another is just barely sufficient to be named partial. Thus, another visual inspection of the timelines was conducted, this time to search for similarities/differences between adjacent groups. The goal was to generate meaningful categories that could be used to distinguish between students whose understanding of large numbers is or is not sufficient to comprehend deep time.

Category	Description	Number of students
7	All 3 TL's correct	10
6	TL 1 & 2 correct; TL 4 partial	6
5	TL 1 correct, TL 2 partial, TL 4 incorrect	2
4	TL 1 correct, TL 2 & 4 incorrect	3
3	TL 1 & 4 partial, TL 2 incorrect	1
2	TL 1 partial, TL 2 & 4 incorrect	5
1	All 3 incorrect	8

Table 4.9 Categories of students by scores on timelines (N=35)

Timelines of students in categories 7 and 6 were very similar to one another. Everyone in both groups was clearly correct on Timelines 1 and 2. None were borderline cases. The only distinguishing difference between the groups on TL4 is that those in category 6 lost proportionality with 100,000, 1 million, and 100 million. While it is possible that the two groups are indeed different from one another, there are other viable explanations for the difference. First, the timelines were drawn fairly late in the interview, and this was the fourth one. Some students may simply have been tired of drawing timelines. Thus their final attempt was not indicative of their understanding. The second possibility is that the cognitive demands of sequencing seven numbers versus four were too great for people in category 6. This does not automatically mean their understanding of large numbers is poor. It may simply mean that seven numbers unnecessarily complicated the task for them. Therefore, these two categories were collapsed into one entitled: *possessing sufficient knowledge of large numbers to deal with deep time.*

Students in categories 5, 4, 3, and 2 were also quite similar. This may seem surprising to conclude that four separate categories could be viewed as similar to one another, yet there are reasons why that is the case. Individuals scored as partially correct on one of the timelines just missed scoring correct, while those scoring correct on a timeline barely made the criteria. Their timelines were more similar than they were different. One of the two individuals in category 5 was moved to correct on TL1 based upon his audio explanation. There is one student, Kayla (8), whose placement in this category may be generous. She will be described more fully in section 4.2.3. Whether she is placed in this category or the next one does not affect a conclusion about her ability to deal with the range of numbers necessary to understand deep time. These four categories were collapsed into one entitled: *possessing insufficient knowledge of large numbers to deal with deep time.*

The final category was comprised of the eight students who scored incorrect on all three timelines. Their first timelines were sufficiently different from those in all other categories that they could not reasonably be combined with another group.

These people were unable to deal with proportional relationships among relatively small numbers, never mind large ones. They were placed into a category entitled: *possessing a poor understanding of smaller numbers.* Categories and the number of students in each category are reported in Table 4.10. Data are also broken out by grade levels. Students in each of these groups share characteristics in terms of how they see the numbers involved in the timelines and in the problem solving strategies they employed to complete the tasks. I now consider each of those groups in turn.

Category	**Number of students**		
	8th grade	11th grade	university
Possessing sufficient knowledge of large numbers to deal with deep time	2	6	8
Possessing insufficient knowledge of large numbers to deal with deep time	5	2	4
Possessing a poor understanding of smaller numbers	5	3	0

Table 4.10 Student groupings based upon their understanding of large numbers (N=35)

4.2.2 Students who possess a poor understanding of smaller numbers

Five of the eight students who scored incorrect on all three timelines were eighth graders and the remaining three were eleventh graders. Their timelines share several characteristics. These individuals seemed to be reasoning from left to right by starting with the smallest number then working up, a strategy that makes it difficult to place the final number proportionally. For several students, the year served as a reference point or benchmark on TL1. They emphasized the need to distinguish a year from a day and a month because a year was a long period of time. Jenna (8) and

Jamal (8) both displayed this thinking. Jenna placed a year at slightly less than the halfway point of her timeline. When asked why she put it there she said that she didn't want to make the space too small because then it would seem like a year is really small. Jamal said,

I did the year a little longer because a year; it's a long period of time. (Jamal, 8)

Their strategy resulted in the creation of a scale that is not clearly logarithmic or linear. Often they compared two numbers at a time rather than trying to conceive of a scale that could handle all four numbers simultaneously, in the case of the first two timelines. The end result was that they changed the scale between each successive pair of numbers. They kept the size of the unit the same between successive pairs, (Ben used the term "tick marks") but they changed its value. Two students specifically stated that the space between a month and a year on the first timeline should be smaller than the space between a day and a month. Ben's explanation makes this point quite clearly.

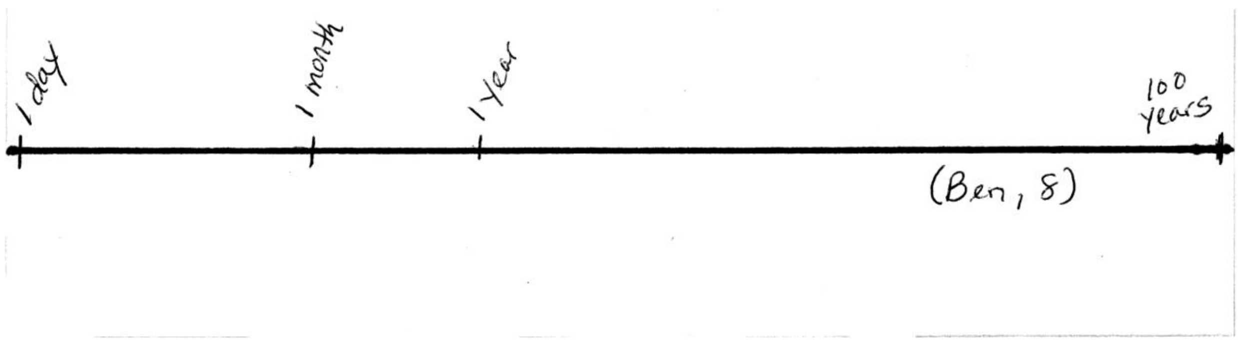


Figure 4.1 Ben's Timeline 1

Ben: A month's like 30-31 days so I put 31 tick marks. I left room for 12 tick marks because there's 12 months in a year. Then here 100 tick marks in 100 years.

Interviewer: Would all your tick marks be the same size?

Ben: Yeah, they would be the same size and spread out evenly.

Interviewer: Even though a month is longer than a day?

Ben: Yeah

Leah used similar reasoning although her actual timeline made the distinction less clear than Ben's did.

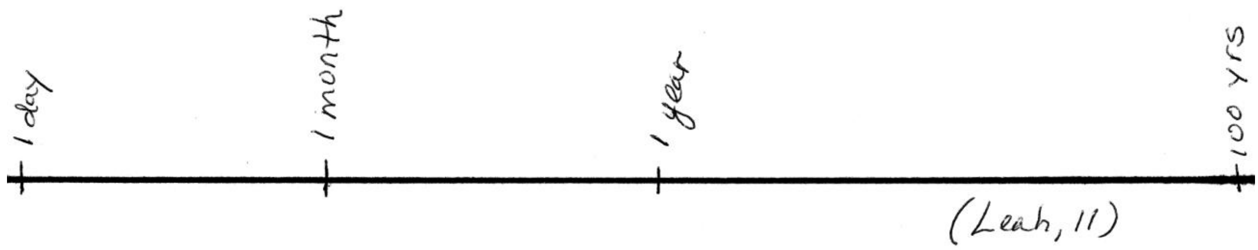


Figure 4.2 Leah's Timeline 1

*I actually said that a day might actually be more space between a day and a month than between a month and a year because it takes [an] average [of] 31 days to get a full month to get to one month. Then it takes 12 months to get to one year. But then it takes 100 years obviously to get to 100 years. So maybe this one then [*points to space between month and year*] would be the smallest one.* (Leah, 11)

Leah continued with this same type of reasoning for TL2. She said she tried to make the space between 1 million and 100 million the same size as the one between 1,000 and 100,000 because there are 100 between both of them, while the space between 100,000 and 1 million should be smaller because there's only ten between those numbers. Malik (11) used similar reasoning and created the following timeline.

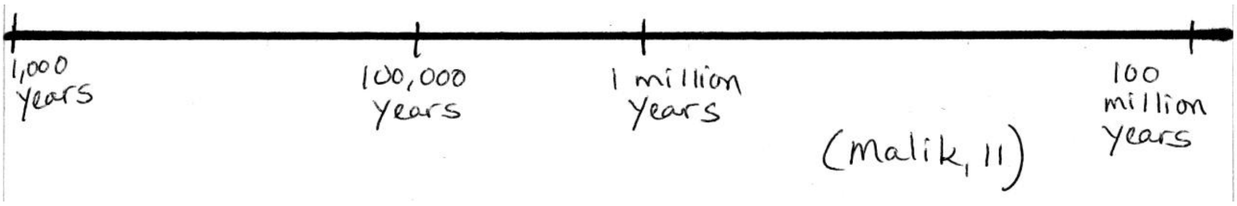


Figure 4.3 Malik's Timeline 2

Interviewer: it seems like you're thinking about something.

Malik: The 100,000 years to 1 million if there's 100 in 1 million years. I'm not sure now.

Interviewer: Hmm

Malik: But that's what I was trying to do, make it equal.

Interviewer: So how many hundred thousands do you think there are in a million?

Malik: I think it's only 10, yeah, I think it's only 10.

Interviewer: Does that change your timeline?

Malik: Yes, I would make this smaller [space between 100,000 and 1 million] because it's the only one of the amount divisible by 10. The others are divisible by 100.

Ben referred to "tick marks" for TL2 as he had done for TL1, but he was also confused about the multiplicative relationships between adjacent pairs of numbers. When asked why he spread the numbers out evenly across his timeline, he paused and then said,

Actually, I think you would put 10 tick marks between each one because I think it's times ten for each one.

Interviewer: So a million times ten to get 100 million?

Ben: Uh, I don't know. I kind of guessed on that one.

Jamal (8) described an additive rather than a multiplicative strategy to compare adjacent numbers. When comparing 100 thousand years to 1 million he said it takes nine hundred thousand years to make 1 million years. He seemed to be thinking that if he has 100 thousand years he needs to add 900 more of those to make 1 million rather than thinking that 1 million is ten times greater than 100 thousand. While Jamal was the only one in this group to employ an additive strategy, it was used by several others in another group and will be discussed more fully in the next section.

Some students in this group indicated they understood little about the numbers involved on Timelines 2 and 4—a more basic problem than the multiplicative relationships among the numbers. One thousand and 100,000 were written as Arabic numerals on the stimulus cards for TL2 while 1 million and 100 million were written using the word million. Vanessa (11) put 100,000 & 1 million at the same point on the timeline because, *"they're the same thing."*

TL4 proved to be an impossible challenge for this group. Students commonly spaced the letters on TL4 evenly across the timeline as can be seen from Jenna's effort. Despite the fact that they had used the entire space available for Timelines 1 and 2, three students used only part of the paper for TL4. This could be due to the two factors that were mentioned earlier. Perhaps the cognitive demands of the task were too high for these pupils. Alternatively, they may simply have been tired of creating timelines.

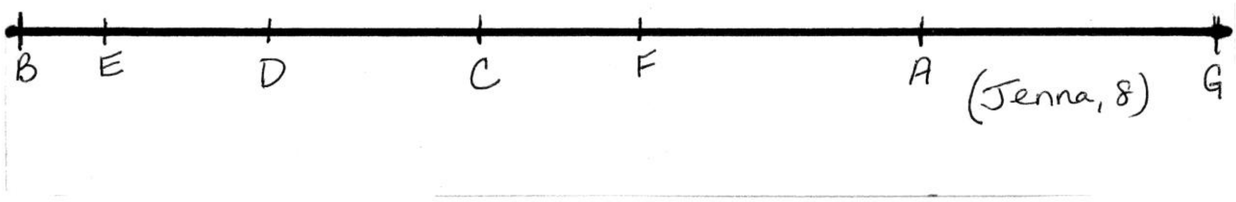


Figure 4.4 Jenna's Timeline 4

4.2.3 Students whose understanding of large numbers is insufficient to deal with deep time

This group could best be characterised by a "great variety in responses and strategies." Six of these students created a partially correct TL1, and the remaining five drew a correct one. Those who produced a partially correct TL1 were more likely to place the smallest and largest numbers first and then place the middle terms. This contrasts with the students in the previous group who employed a left to right, smallest to largest, orientation for all three timelines. Students whose scales were more logarithmic were concerned about showing the differences between the three smallest numbers. This resulted in the spacing between the final two numbers being too small, as is shown by Ashley's (8) TL1. This could indicate that they are dealing with adjacent pairs of numbers separately and are not considering them all on the same scale.

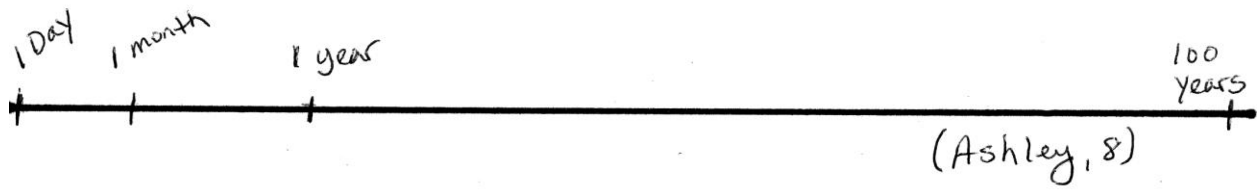


Figure 4.5 Ashley's Timeline 1

Some timelines in this group could best be described as borderline and were only placed in the partial or correct category based upon their verbal explanations. Kayla (8), who was mentioned in section 4.2.1, is a student whose TL1 was moved to the partial category based upon her explanation. The timeline she produced is less proportional than Ashley's and is more similar to Leah's. Visual inspection alone placed her in the incorrect category. When explaining her timeline Kayla appeared to be starting from one day as her basic unit and referencing the other times to that. She mentioned 30 days in a month and 365 days in a year. She said, "Then 100 years, that's a lot of days." When it was pointed out to her that the spaces between a month and a year and a year and 100 years were very similar on her timeline, she said that her timeline wasn't long enough. Kayla then indicated a space about 15 centimetres beyond the right side of the timeline as the place where she would locate 100 years if the paper was longer. Although her strategy was similar to those in the previous group, that statement suggested to the interviewer that she had a clearer concept of the relationships among the numbers than what was indicated by her timeline alone. Therefore, she was given the benefit of the doubt and moved to the partial category. In questionable situations throughout the thesis, I have chosen to err on the side that would under rather than over report difficulties in any of the three areas being investigated.

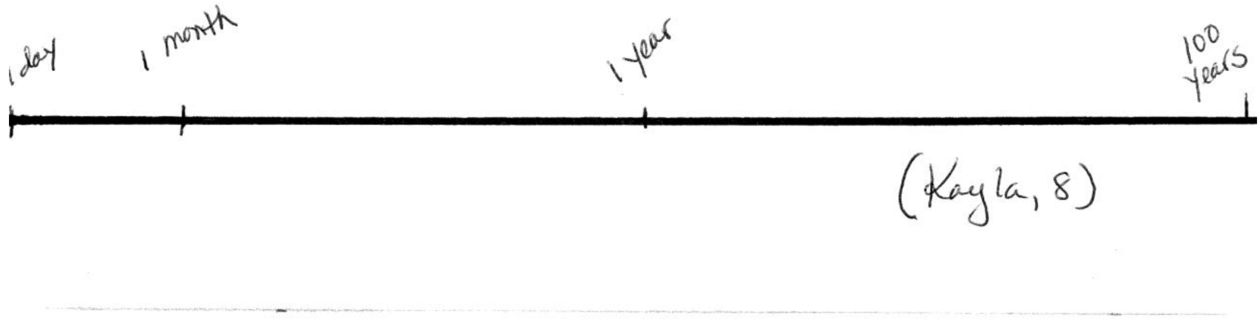


Figure 4.6 Kayla's Timeline 1

Nicole (univ) is another example of a student for whom a timeline was moved to a different category based upon her explanation. Her TL1 was initially categorised as partial and is very similar to Ashley's. The only real difference is the small arrow by 100 years. Her reasoning strategy was reminiscent of Kayla's described above. Nicole talked about reasoning from one day to get to a month, then multiplying 30 times 12 to get one year, and then multiplying 30 times 12 times 100 to get 100 years. When questioned about the small arrow by 100 years, she indicated a space about 37 centimetres beyond the right side of the timeline where she would place 100 years if the paper was long enough. At that distance she would have created a timeline that met the criteria set out in Table 4.7.

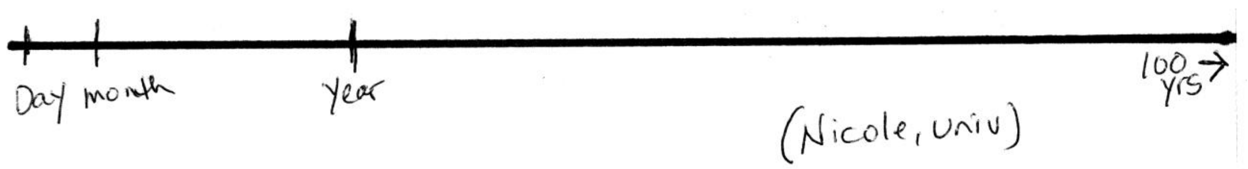


Figure 4.7 Nicole's Timeline 1

What was most striking about the students in this group was the marked difference between Timelines 1 and 2. All of their TL1s showed some evidence of trying to distinguish the relationships between adjacent pairs of numbers. However, on TL2 they frequently spaced all numbers equally across the line. Of the eleven students in this group, nine of them scored incorrect on TL2, including three students who scored correct on TL1. As was typical of those in the poor category, many students expressed confusion about the relationships between adjacent numbers in their explanations. Several thought all were multiplied by 100 to get the next number, while others thought all were multiplied by 10. A number of them mentioned the difficulty of the task. Danielle's (univ) TL2 was quite similar to Leah's TL1. She spaced the numbers out evenly across the timeline with a slightly larger space between 1 and 100 million than the others. As she was working she said, *"The numbers between 100,000 and 1 million are very blurry."*

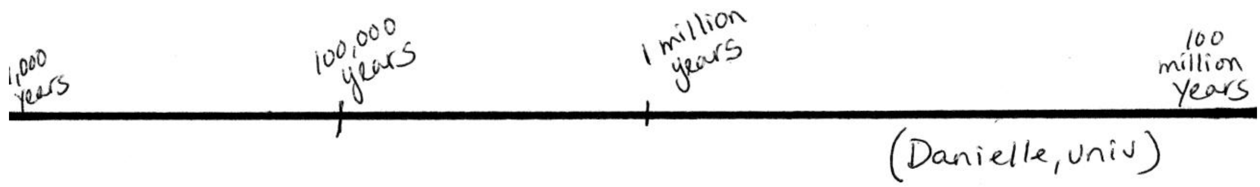


Figure 4.8 Danielle's Timeline 2

Emma (8) mentioned similar confusion about the numbers involved.

Interviewer: You made the distance from 1 year to 10,000 years about the same as the distance from 10,000 years to 10 million years. Are they the same amount of time?

Emma (8): No, one's longer than the other.

Interviewer: So why did you make them the same distance? I'm curious about that.

Emma: I don't know. I guess I wasn't sure where to put them. …It's a long period of time so I wasn't really sure where to put it.

When asked how to explain the relationship between two of the numbers, two students explicitly used additive rather than multiplicative reasoning, comparable to what was described for Jamal in the previous section. The timelines they drew were typical of those in this category with the four numbers spaced fairly evenly across the timeline, similar to Danielle's. Jamal didn't use the word "more" in his explanation, but Kayla (8) did when she indicated it takes ninety-thousand more thousands to make 100,000, nine hundred thousand more hundred thousands to make 1 million, and ninety-nine more millions to make 100 million.   Emma also used this strategy. She subtracted correctly on her timeline but then misspoke when giving the relationship between 100,000 and 1,000,000.

Interviewer: How many hundred thousand years does it take to make 1 million years?

Emma (11): Probably [pause], can I write this down?

Interviewer: If you want, sure. [She subtracted 100,000 from 1,000,000 on the back of her timeline and got 900,000.]

Emma: Probably about 900 years.

As was true for the previous group, the need to place seven numbers on the same timeline that represent durations that are vastly different from one another created problems for many on TL4. As a result, in some cases the student's response didn't make sense. Ashley, another eighth grader, illustrates.

Interviewer: You have about the same amount of space between the year and 10 thousand years as you have between 10,000 years and 10 million years.

Ashley (8): Yeah, cause they're like be [sic] the same amount.

Interviewer: From a year to 10,000 years would be the same as from 10,000 years to 10 million years?

Ashley: Yeah, cause they're just another year or so

The use of an intuitive logarithmic scale shows up in several places on TL4. Two university students demonstrate their confusion.

Elizabeth (univ): Well, I thought that it would probably take more minutes in a day, 24 x 60, I don't know how much that is, but it would probably take more minutes to make 1 day than it would take 1 day to make 1 month. A minute is like a smaller amount of time so I kind of thought they would be about the same amount of time.

Leah (11): Then for these [10,000, 10 million, & 100 million], they feel like they're the same space to me. They just feel like they're the same to me.

Interviewer: The same distance apart? You mean the same distance from 10,000 to 10 million as from 10 million to 100 million?

Leah: Well, I'm not sure what I'm saying but that's what I was trying to do.

Nicole's (univ) reasoning was more explicit.

Nicole: I knew that the minute was the shortest so I put it at the beginning of the timeline and I knew that 100 million years was the longest so I put it at the end. And I tried to fill in the spaces between that. But the way that I did it for this, because if I would have done it by the minute scale, I kind of like changed my thought processes for each one. I did 60 minutes in an hour, 24 hours in a day, and I tried to do this in terms of minutes. Then for the day I did 30 days into a month and then 12 months into a year. Then, you know, 10,000 lines right here for 10,000 years, because the timeline wouldn't be long enough if I had to do it like how I thought it would be. Then for this one, I know that there's a lot of notches that would go from 10,000 years to 10 million, definitely a lot for 10 million to 100 million if I did it on a minute scale.

Interviewer: So you changed your scale between each two amounts?

Nicole: Yeah, to make it fit on the paper.

What these students seem to have in common is the attempt to clearly differentiate between 1 year and 10,000 years and sometimes 10 thousand and 10 million years. The result is that the spacing for the remaining numbers is particularly out of proportion. In some cases 100 million was simply put at the end "because it's the largest" or because the biggest number always goes at the end. Other students in this category demonstrated that they understood that on a linear scale the distance between 10,000 [F]and 10 million [A] must be greater than the distance between 1 [C] and 10,000 [F]. Danielle is an example of the latter type, yet it's not clear if she simply ran out of room and quit or if something else is going on. The fact that she felt the space between a month [D] and a year [C] needed to be greater than the space between a day [E] and a month suggests that she does not realize that on this scale the difference is inconsequential.

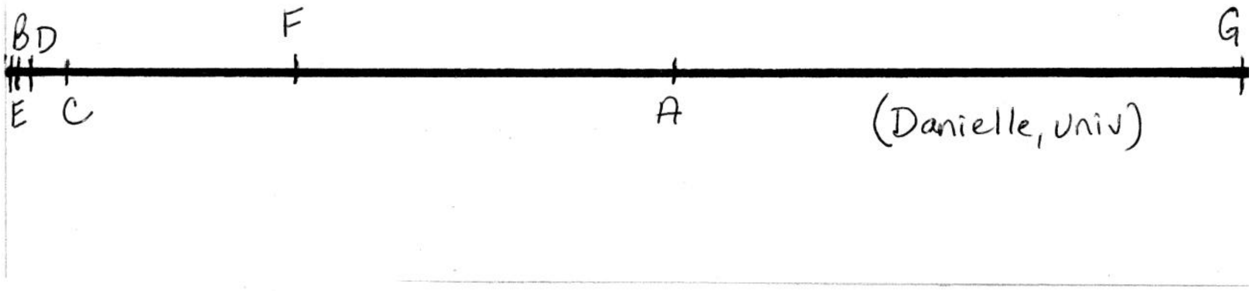


Figure 4.9 Danielle's Timeline 4

Danielle (univ): B is a minute. E is one day. The spaces between those compared to all the rest of them—very little. Then a day compared to a month, there's 30 days about in a month, then there's a little more space. 12 months in a year give it a little bigger. Then trying to think where the difference between 1 year and 10,000 years is probably much greater than that but it can be shown like that I guess. Then what's the difference between 10,000 years and 10 million years? Oh that would be huge, so much bigger. Then your 100 million years would be all the way at the end.

4.2.4 Students whose understanding of large numbers is sufficient to deal with deep time

These timelines differed from the others in that there was a clear recognition that in order to represent all four numbers in the first two timelines on a linear scale, the first three numbers would be very close together at one side of the timeline. Cole's TL2 typifies the group.

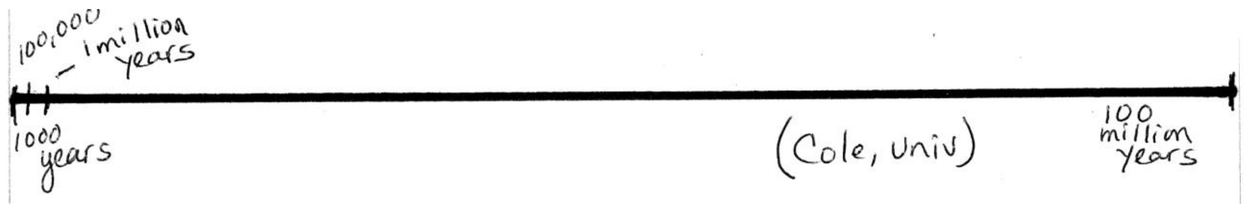


Figure 4.10 Cole's Timeline 2

These students differed from the others not only in their understanding of number, but also in terms of their problem solving strategies. They were much more likely than the other groups to place the largest number first, make several attempts to place the first three numbers, and move them progressively closer to the left side of the line. Justin was an example of someone who made several attempts.

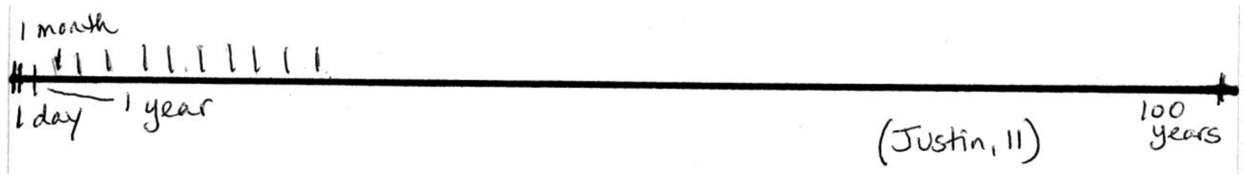


Figure 4.11 Justin's Timeline 1

Well, I would make the increments, the day. I didn't know how far apart to put them. I found out how far apart the one was. I just found out how much I wanted to put one. Then I used my pinkie to see the size and I moved it over and marked it every time until I got to 30, put a line down for a month, and then tried to gauge my pinkie the other way so I could get all 30 onto it and made that as one month and then just kept moving the pinkie over until I got to 12 until I got to a year. That didn't fit at first so I decided to make it even *smaller, still using my fingers to measure it up. By the time I got to the 1 year point I wouldn't be able to fit 100 years so I had to eventually make the year, the day, and the month pretty much in the same general starting area showing that they are so close in comparison to 100 years that it would fit.* (Justin, 11)

Michael (11) described a unique division strategy for TL1. He began by dividing the timeline and writing the numbers 50, 25, 75, and 12.5.

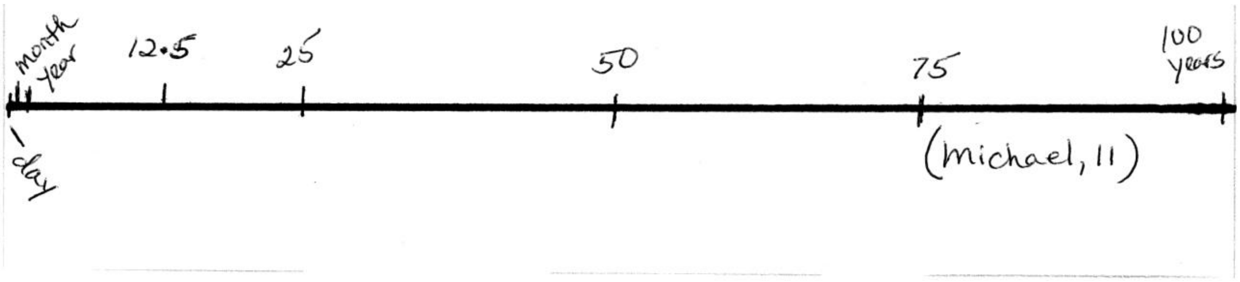


Figure 4.12 Michael's Timeline 1

I split in half 100 years here at the very end, split in half, got 50, split the two ends in half, 75 and 25, 25 in half, 12.5, that in half, 6 point 2 whatever, then I just guessed at one year be about here, and a month should be a little more over [left], and the day's the smallest.

No one in the other two groups mentioned any connection between the timelines but students in this group did connect them. In fact when they began TL2, a number of them commented that it was very similar to the first one in that the first three numbers were clustered to the left side of the line with the fourth number alone at the right side. While completing the first two timelines a few students mentioned the need to reference the three smallest numbers against the largest number.

This group frequently mentioned a multiplication or division strategy in their explanations of their timelines. Ryan articulated that for TL2.

Interviewer: Tell me what you know about how these numbers are related to each other.

Ryan (11): Multiplied by 100 [indicates 1,000]

Interviewer: You mean you multiply 1,000 by 100 to get 100,000?

Ryan: yeah, then you can multiply 100,000 by ten to get a million. Then multiply that by 100 to get 100 million, so all factors of ten. (Should be multiples]

Several students talked about how the first three numbers on TL2 were related to 100 million. While not all students mentioned the specific multiplicative relationship between the numbers, many indicated the relative difference between the numbers. Sarah's (univ.) explanation is an example.

Interviewer: It's interesting because 1,000 years seems like a long time.

Sarah: Yes, it does.

Interviewer: Yet it doesn't take up very much space on your timeline.

Sarah: Not at all

Interviewer: Why is that?

Sarah: Because it's a lot, a lot smaller than a hundred million years.

Mathematical language was used more frequently by these individuals. Four of the students mentioned the word “proportion” or “proportional,” in contrast to the previous two groups, in which it was never mentioned. Those who used the word proportional expressed how difficult it was to make the timeline truly proportional. Sean talked about proportional reasoning for both TLI and TL2.

> *When it comes to what was going through my head, I was thinking math, math, math the whole time. I was thinking proportions. You know, if a day is this small on the chart then it has to be, you have to multiply it by so many to make 100 years. But it’s kind of hard to do without if you want to figure it out mathematically. There’s other ways using a ruler or calculator.* (Sean, 11)

His explanation for TL2 is very similar.

> *Sean: I didn’t do it thinking about the other one. I did it thinking in proportion thinking that since I’m looking at such big numbers I should break it down so I thought ok this is just one instead of 1,000. I thought just one, just 100 for the 100,000. Then just 1,000 for the [million] cause I’m just thinking in the range of the three zeros at the end for the 1,000.*
>
> Interviewer: Do you mean because a million is 1,000 thousands?
>
> *Sean: Yeah, I thought of it [100 million] as 100,000 thousands...It pretty much is like the other one not in number wise because obviously one day compared to a month, a month has 31 days in it, a year has 365 days in it. The last one was 100 years, just 100 years itself so those numbers aren’t all the same as these but they could be linked in proportion.*

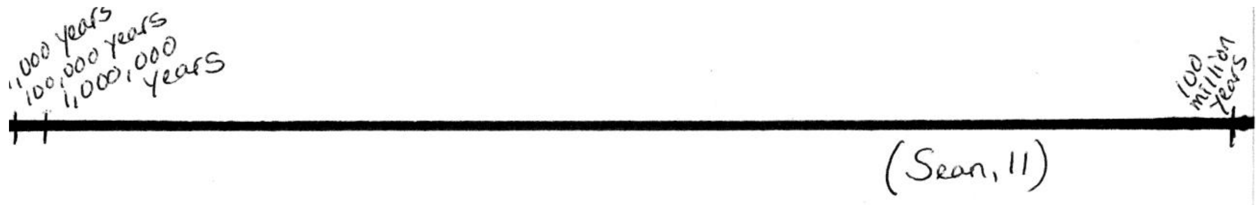


Figure 4.13 Sean's Timeline 2

Although these students created accurate timelines and explained their solutions using some sort of multiplicative or proportional reasoning, they didn't always get the relationship between adjacent numbers correct in their explanations. Initially Justin (11) said there are 100 hundred thousand years in a million years, although immediately afterward he said that he would multiply the space necessary for 100,000 by ten to get the space needed for 1 million. Lauren was not immediately sure of the relationship between 100,000 and 1,000,000 either.

Interviewer: How many hundred thousand years are in a million years?

Lauren (univ): I have no idea, but I know there's not a lot.

Interviewer: Why do you know there's not a lot?

Lauren: Because million there's one more zero so that's like 10,000 more of these. I don't know. My brain can't comprehend it right now. I know it's smaller by quite a few.

Where students in this group diverged from one another was on TL4. Michael said he wanted to indicate that the timeline should be longer. When asked to draw the line as if this was the only paper he had, he produced the timeline in Figure 4.14. He tried to indicate the great difference between a year and 10,000 years, and 10,000 and 10 million years. Neither his timeline nor his explanation provided enough information to say whether he wasn't aware that his spacing between 10 million and 100 million years was entirely too small on that scale or if he simply chose to leave it as "good enough."

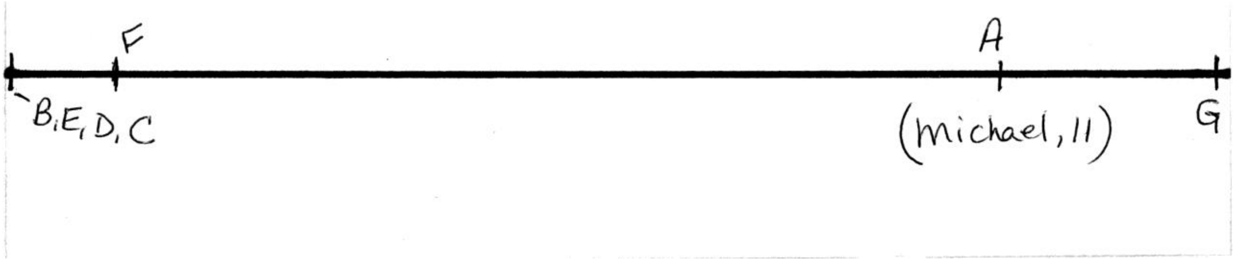


Figure 4.14 Michael's Timeline 4

Lauren (univ.) attempted proportionality with the smaller numbers up to 100 years, although one could argue that on a scale of 100 million years, they were still quite far apart. Yet, her space between 10 million years [A] and 100 million years [G] was smaller than the space between 1 year (C) and 10 million years (A). Once again, this could suggest that the cognitive demands of placing seven very different numbers on the same timeline were too great for some individuals.

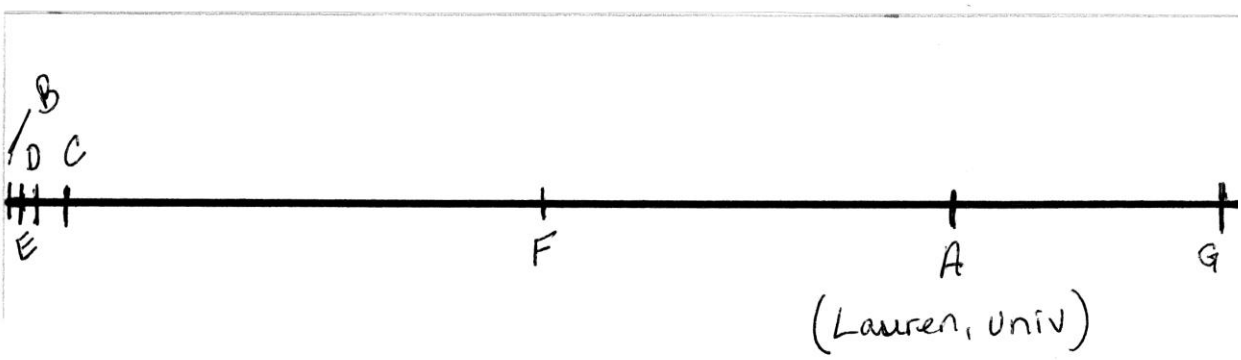


Figure 4.15 Lauren's Timeline 4

Other students in this category appeared to approach TL4 in the same way they had the earlier ones, with many once again mentioning a multiplication or division strategy. They also appeared to have a clearer sense of the scale required to deal with all the numbers, although in some instances it was clearer than others. Anthony specifically said that all the other numbers must be considered in terms of their relationship to 100 million years.

> *Anthony (univ): B, C, D & E were all within a span of a year and in geologic time compared to 100 million years that's just like a decimal percentage so that's towards the beginning. F just cause you're adding 10,000 years. It might have been a little bit bigger but not too much more different. Then 10 million years would be a tenth of this since I decided to put 100 million years so I just kind of estimated that'd be 10% of the graph. Then G would be what you're basing everything off of.*

James also demonstrated a sense of scale. He said that it would be difficult to show all the numbers because the first few are so small in comparison to 100 million years. Notice that he indicated that B [1 minute] should be against the edge of the paper. In fact, he said, *"I ran out of room. B just went against the line cause I couldn't draw any more lines."*

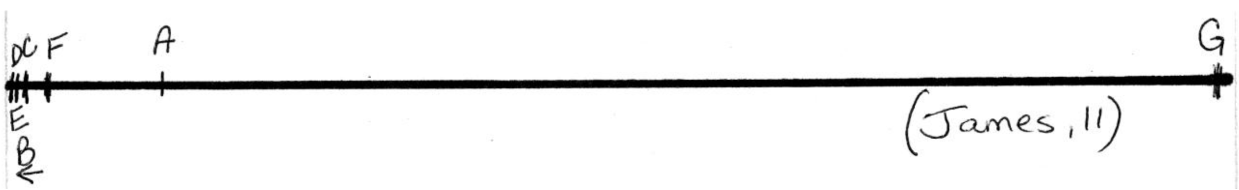


Figure 4.16 James' Timeline 4

Although he didn't state it as clearly as Andrew, Ryan drew his timeline to show that the smaller numbers are essentially the same when thinking on a scale of 100 million years.

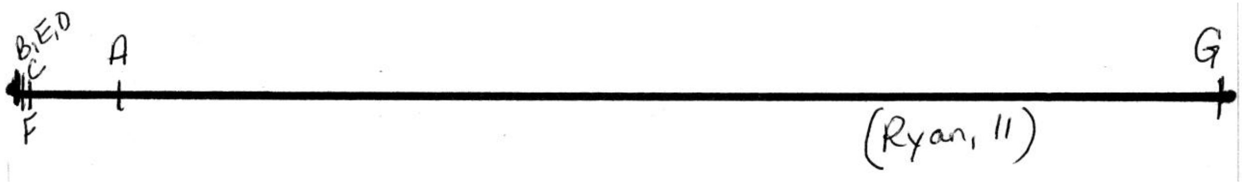


Figure 4.17 Ryan's Timeline 4

Ryan (11): In 100 million years you have, what's that, 10 million years, you have 1/10 of that. That's what I think is 1/10 of the entire timeline. The rest of them are almost nonexistent. If I could have squeezed it closer I would have because 10,000 years is 1/1000 of this so that would be all the way at the end.

4.2.5 Anomalous data

I have already mentioned why some timelines were difficult to categorise and have attempted to justify the reasons for their placement in a particular group. One student's (David, univ) responses were the most difficult to classify. He is indicated by an asterisk in Tables 4.7 and 4.8. His responses were so unique that I have chosen to discuss them at length, partly to illustrate the challenges inherent in interpreting interview responses. David was the only university student to identify himself as a geology major, though like all university students in the sample this was his first university level geoscience course. Initial sorting by both coders placed all three of his timelines in the incorrect category. All are reproduced for the reader's convenience. Each was difficult to interpret even with his explanation as it was unclear if, for example one year was deemed to be a fixed point on the timeline or a line segment indicating its duration.

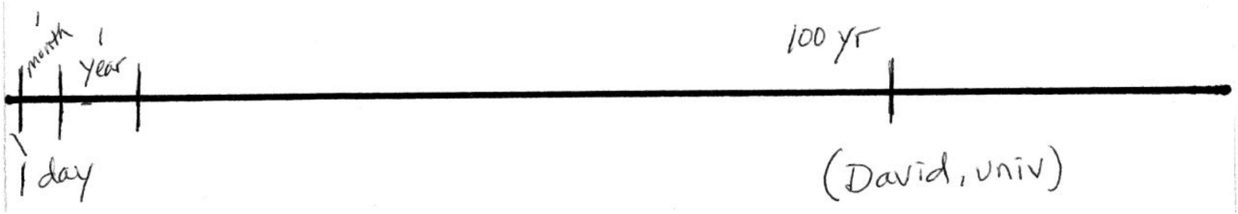


Figure 4.18 David's Timeline 1

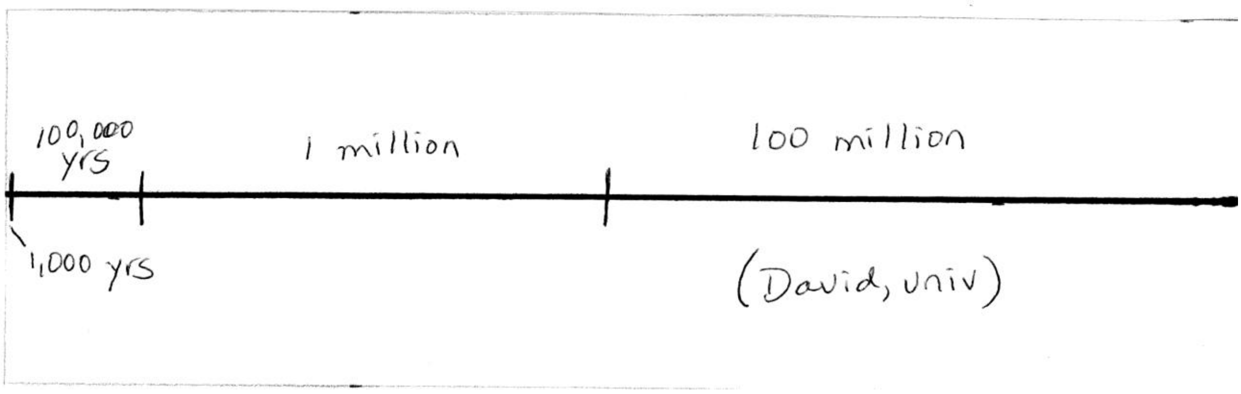


Figure 4.19 David's Timeline 2

After listening to the audio tape for TL1 and TL2, he was moved to the partial category. When creating TL1 he made multiple corrections and finally said, "It's not proportional but it's closer than what I had." As he worked on TL2, he said,

> *I'm just thinking of how big these numbers are compared to each other. I'm not quite sure 'cause we don't deal with numbers this big normally in our average day. So I really don't know how long one million years is compared to a thousand. I really don't know how long a thousand years is. It's kind of hard to put that in the reference when I don't know how long a thousand years is*

much less a hundred million years. It's kind of hard to put that on the timeline. I'll do my best but I don't know so I'm just gonna keep going. (David, univ)

That explanation suggested an intermediate understanding of the numbers. Yet his rationale for TL4 implied a much richer understanding of scale than either of his previous two explanations. He expressed confusion about the size of the numbers similar to what he said for TL2, but this time he indicated that the problem was a more general one of scale by relating these times to very large objects. The fact that he was able to relate large units of time to large size units suggests that he is not necessarily saying that he doesn't know the proportional relationships among the numbers, although that may be the case. When explaining his placement of numbers for TL2 he said that each successive number was 100 times greater than the previous one. It is unclear if that represents a true misunderstanding or whether he was simply speaking without fully thinking through his response. When talking about TL4 he appeared to be saying he has difficulty fathoming such large periods of time rather than simply ascertaining the proportional relationship between two quantities.

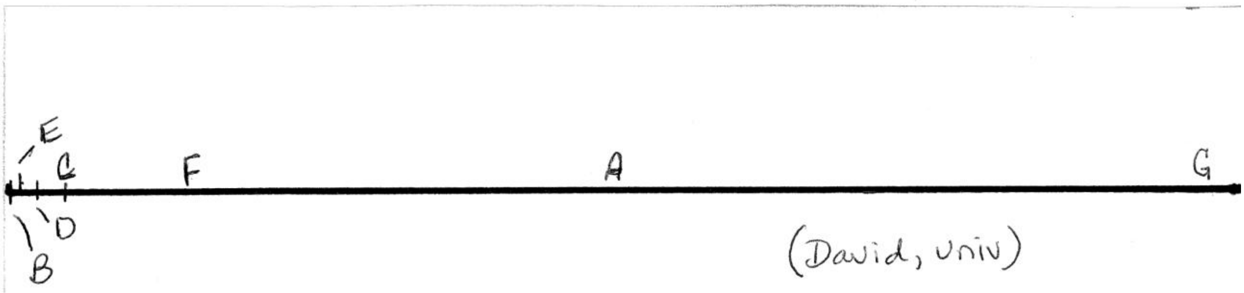


Figure 4.20 David's Timeline 4

David (univ): 100 million years takes a lot of minutes, just the scale, just the way, I mean, I don't really know how to explain it other than…umm

Interviewer: Do you mean you started at the right side of the timeline in your thinking?

David: Not really. I started down here but I kinda did start all the way at the right. What I did was I took a look at all of these, A through G, and I saw the shortest one is a minute and the longest is 100 million years. That's gonna be a large time scale. And I looked at all of what is in here, in between a minute and 100 million years, 1 year is not that long. 100,000 years isn't even that long. A day obviously isn't long, a month is not that long either. That's why I bunched them down here. Cause 100 million years is a long, long time. That's pretty obvious, but that is a long time. It's kinda like when you get up to excessively high temperatures or excessively big objects such as stars or planets. You can't even fathom how big those things are. They're just so big; I don't know how big a star is. It's just so big, I don't know. And that's the same way with the years; it's just so long that I have no clue. Like I can judge a minute. I can judge a year. I can even judge 6 years, but after about 6, 7, 10 years it's kind of hard to fathom, and that's just 20 years, 10 years. Imagine 100 million years? I can't do that, no.

David's number lines suggest a fairly sophisticated understanding since he appears to be thinking beyond a mathematical relationship to the need to make sense of the numbers which is a great challenge for anyone attempting to comprehend deep time. Therefore, he appears as correct for all three timelines in Tables 4.7 and 4.8 with an asterisk. Again, in an attempt to resist overstating the role of a large number problem in a conception of deep time, it is deemed preferable to possibly introduce a systematic error that underestimates rather than overestimates the difficulty.

4.2.6 What do the timelines indicate these students understand about large numbers?

The second research question asked whether students in this sample possess sufficient understanding of large numbers to be able to comprehend deep time. Results demonstrate that a number of these individuals have relatively inadequate ideas about numbers of the magnitude that are necessary to understand deep time. All eight students who were classified as having a poor understanding of smaller numbers demonstrated problems with multiplicative reasoning as well as a more basic misunderstanding regarding the meaning of the specific numbers in the task. They appeared to approach all three timelines in a similar manner with a left to right strategy that compared two numbers at a time rather than looking at the set of numbers as a whole. These students also tended to change the size of the unit represented by the spaces between successive numbers. In other words, they talked about the need to fit 30 spaces between a day and a month and then the need to fit 12 of the *same size spaces* between a month and a year. They perceived numbers that were relatively small on the scale as being large. As Jamal (8) said, *I did the year a little longer because a year, it's a long period of time.* Students in this group rarely moved any of their numbers while explaining their reasoning. As expected they knew how many days in a month, months in a year, etc. They were not so clear on the relationships between numbers for the other two timelines, however. Some thought adjacent numbers were related multiplicatively by 10 and others by 100. They would not be expected to deal very effectively with long durations.

Students in the second group were able to deal with relatively small numbers, but their Timelines 2 and 4 were not appreciably different from those of the prior group. They were also confused about the multiplicative relationships between pairs of numbers. They tended to assume the same multiplicative relationship between each adjacent pair of numbers. Hannah (11) said that the space between 100 thousand and 1 million should be larger than the spaces between 1,000 and 100,000 and between 1 million and 100 million. Her reasoning was that 500 [thousand] is half of a million and 100,000 isn't even at the halfway point. When asked how many 100 thousands in 1 million she replied, "A lot, I'm not sure exactly."

Two students in this group and one in the previous group attempted to use an additive strategy rather than a multiplicative one to relate two numbers to one other. On the first two timelines, people in this group often placed the smallest and largest numbers first and then placed the middle two numbers in relation to them. They tried to clearly show the distinction between smaller numbers like 1,000 and 100,000 (TL2) or 1 year and 10,000 years (TL4). The end result was that they ran out of room and telescoped the space between the two largest numbers on the line. They struggled with TL2 but TL4 was clearly too great a challenge. The numbers themselves, apart from any difficulties with geoscience content would likely impede the ability of these first two groups to deal with either succession or duration in deep time.

The final group demonstrated a solid understanding of the relationships amongst the numbers in this task. They displayed good proportional reasoning. In fact, the way they reasoned about the task set them apart from the other two groups. They kept the scale in mind and were able to consider the numbers as a group when deciding their placement on that scale. A number of students in the correct category noted that each timeline was essentially the same task just with different numbers. Talking about TL2,

> *It's almost the same principle.* (Cole, univ)
>
> *Well, it's kind of the same idea as the day and the month.* (Sarah, univ)
>
> *I think it's like the same thing.* (Lauren, univ)

They tended to start with the largest number and relate all the others to it rather than the strategies used by the other groups. They recognized the role played by the largest number in the set on the scale. As a group, they were more likely than the others to use the word proportion, proportional, or fraction in their explanations. It was not unusual for these participants to change the placement of numbers on their timelines as they worked. If these students have any difficulty understanding deep time, it will probably not be due to the large numbers involved.

There appears to be a slight age effect for the timelines (see Table 4.10, p. 240). However, in positing an age effect, it must be pointed out that the university sample is not stratified like the eighth and eleventh grade ones. On the other hand, only one of the university students self-identified as a science major. The other eleven participants were enrolled in the course to fulfil a university general education requirement. Further, demographic information for Institutions A and B reported in chapter three indicates that students at these schools are representative of students within the average to below average range of ability for all university students in the U.S. Hence, there is no implicit assumption that the university students in this sample represent a population that is more academically talented than the U.S. university population as a whole. While I would not be on firm ground to allege too strong an age effect, there is a hint that one exists.

We see, then, that some adolescents and adults do not possess a solid understanding of large numbers. They could be expected to have difficulty comprehending deep time since they cannot relate time units to one another on a suitable scale. Simply learning about geoscience will not be enough. They will also need to learn something more about large numbers and their proportional relationships. There is another factor that may impede a student's ability to understand deep time. We now turn to the final "leg" of the "stool."

4.3 Geoscience content knowledge and a concept of deep time

The role of geoscience content knowledge in students' understanding of deep time seems obvious, yet it is difficult to isolate. From a constructivist perspective, all knowledge acquisition is mediated by existing knowledge. In this section, however, we are specifically investigating the extent to which errors students make about deep time can be at least partially attributed to a lack of subject matter knowledge of geoscience.

If a lack of geoscience content knowledge contributes to a poor understanding of deep time, how would we expect that to manifest itself? On succession tasks, we would expect students to be able to order events correctly when the ordering is possible via logical reasoning that is largely independent of their knowledge of geoscience. For example, appearance of first life is placed before appearance of first fish since fish are alive. A student does not necessarily have to know anything about evolutionary biology, but could reason that since fish are subsumed under the more general category of life, first life is antecedent to first fish. Conversely, where subject matter knowledge is more necessary for accurate placement we would expect more errors. Regardless of whether answers were correct or not, we would expect the rationale for answers to focus on more extraneous surface features, e.g., size, appearance of picture. As a result, a correct answer alone would not necessarily be indicative of a good understanding of time since a student could get the correct answer using faulty reasoning. We would also expect that if specific subject matter knowledge was required, student performance would be no better for succession or duration tasks in conventional time. Thus, if a student is asked to indicate how long it takes one blood cell to travel through the body, it would not be surprising if the question proved quite difficult. Some knowledge of the circulatory system would be essential. The fact that the process occurs well within conventional time makes it neither easier nor harder than processes that occur in deep time. The important thing is that one must know something about the domain in question. We first turn our attention to geoscience content knowledge as it relates to succession.

4.3.1 Geoscience content knowledge and fossil succession

Students were shown a line drawing of a series of stratigraphic columns containing specific fossils. This drawing was taken directly from Dodick and Orion (Dodick & Orion, 2003a, 2003b) and was described in chapter two (Figure 2.3 on p.

132). Students were asked to compare two layers, the trilobite and the brachiopod layer and indicate which layer probably formed first or if they formed at the same time. They were deliberately not directed to the locations of the layers as that could have suggested a particular strategy. Principles of stratigraphic correlation and fossil succession can be used to determine that the trilobite layer was formed before the brachiopod layer, and, in fact, was the first one in the sequence. The brachiopod was chosen because it appears in two places in the sequence, one on the same level with the trilobite and one above it. It was hypothesized that this dual placement might cause some cognitive conflict in students. While the presence of the brachiopod on the same layer as the trilobite might suggest the two layers had formed at the same time, the presence of the brachiopod farther up the sequence in column two could challenge that thinking.

Table 4.11 lists the number of students who chose each fossil and the reasons they cited for their answer. Fifteen students said the trilobite formed first, six said the brachiopod did, and 13 said the two layers probably formed at the same time. This total equals 34 not 35 due to the fact that Claire did not settle on an answer but named trilobite first or at the same time as equal possibilities. She cites her lack of knowledge about the two fossils as an explanation for her difficulty. Since Claire did not settle on a response, her reasons are not reported in this table. Further, the total frequency in Table 4.11 does not equal 34 because some students invoked more than one reason for their answer, all of which have been reported. From those numbers alone we might conclude that approximately 43% of the students had a good understanding of fossil succession. Yet, when we probe the stated reasons for their responses, that conclusion becomes more tenuous.

Reason	Frequency		
	Trilobite	Brachiopod	Same
Physical appearance of fossils (e.g., one looks older, one is larger)	4	6	1
Same position in column (both at the bottom)	0	0	12
Greater familiarity with (knowledge about) one of the two fossils	4	1	0
Reference to others in column (more recent, more familiar, name, where lived)	6	1	1
Correlating layers	4	0	0

Table 4.11 Reasons cited for answer to question, "Which is older, the trilobite or the brachiopod?"

Students who answered incorrectly and some who answered correctly cited some aspect of the physical appearance of the fossils, a surface feature, as reason for their answer. None of the university students said this, but it was common among younger participants. These responses included statements about size as well as other aspects of physical appearance.

> *James (11): It [trilobite] looks like it has bones or something and the brachiopod doesn't maybe. It [trilobite] took longer.*

> *Ben (8): It [brachiopod] looks like it's older and everything. The trilobite one looks like it's newer and there's not that much dust and everything on it, dirt.*

> *Kayla (8): I think the trilobite would form first cause it looks bigger than the other one.*

The same position in the column was cited by all but one person who said the layers formed at the same time. There is some correlational reasoning here, since same place in a column equals simultaneity. The problem is that these individuals failed to account for the brachiopod's appearance in two different places. It is not clear, however, if these people ignored the relevance of the brachiopod in the second column or if they didn't see it.

Alyssa (8): I'd say they both formed at the same time.

Interviewer: Why do you say that?

Alyssa: Because they're both on the bottom of the rocks.

Familiarity or lack thereof with one or more of the fossils shows up in two places in Table 4.11, both in the row of that name and also in the reference to other fossils in the same column. The greater familiarity row is reserved for statements indicating students were more familiar with the trilobite or the brachiopod specifically. In general, students in this category often said that the brachiopod looked like a clam, and they were familiar with clams. The reference to others includes statements students made regarding any of the other fossils in the columns. Some mentioned that the fossils in one column were fossils they had heard of before. Therefore, the fossil at the bottom of that column must be younger than a fossil at the bottom of another column. This, however, may not be the case since an unconformity could have removed intervening layers. Matt's response is an example of one that was classified as familiarity/unfamiliarity.

I've never seen anything like that [trilobite] before. (Matt, 8)

Nathan's, Leah's and Sarah's, however, were classified as "reference to others in column." While they cited familiarity with certain fossils, the focus was on the other fossils in the column, not the two being compared.

> Interviewer: Why do you say the trilobite?
>
> *Nathan (11): Um, these [1st column] seem to be like animals, I guess, yeah, animals that seem to be like more older like these [3rd column] are things that are like things that are alive now like snail and shark tooth where these [1st column] like gastropod, I've never really heard of that so maybe it's prehistoric.*
>
> *Leah (11): I was just looking at the ones that had pod at the end were higher up than the ones that didn't.*
>
> *Sarah (univ): I think that the trilobite layer was formed first.*
>
> Interviewer: Why do you say that?
>
> *Sarah: Because the brachiopod layer has things on top of it that are more recent than the things in the trilobite column.*

The majority of people who invoked one of those reasons correctly answered that the trilobite was formed first. The results provide a good example of the fact that it is possible to get the correct answer using a faulty reasoning strategy or one based upon a faulty premise.

It was not unusual for students to combine the physical appearance of the fossils with familiarity with one of the fossils being compared or with others in the columns.

Chris (8): I would say the brachiopod.

Interviewer: Why would you say that?

Chris: It would take less time to encase it cause I've seen some of 'em. I used to collect seashells so I know what it's gonna look like. It [brachiopod] could be possibly the size of that [trilobite] but from the picture it's not so, it's a little smaller surface area so it may take a shorter time to encase.

Only four of the students mentioned correlation of layers, and then with varying levels of sophistication. Justin (11) said the trilobite was first because it was before the ammonite in the first column which was also at the bottom of the second column.

I: Would your answer be different if you noticed there was also a brachiopod here [bottom of 3rd column] or would your answer still be the same?

Justin: It would probably be the same cause then just further down this [3rd column] you could probably find an ammonite; I guess if the ammonites could live in that area then an ammonite would be further down on the third column then even farther than that would be a trilobite.

Connor, an 8th grader, gave the clearest explanation of fossil succession.

Connor (8): I think that the trilobite layer formed first.

Interviewer: Why do you say that?

Connor: Because comparing the brachiopod layer to the ammonite layer, the snail, like taking the snail and measuring up to there, maybe taking the

> *brachiopod and measuring to there. Even though the fish scale is in between those it could be around the same time. Yet, if you take the top of the trilobite layer which is the gastropod, you put it right up over against the gastropod there[in 2nd column] really, it turns out the trilobite's all the way up here and the shark tooth's way up here so that's kinda how I did it.*

If we had merely stopped with the question, "Which layer was formed first?" we may have incorrectly concluded that 15 students had some understanding of succession in deep time. However, when we look at the reasons they cited for their responses we find that roughly the same number of people who chose trilobite (correct) and brachiopod (incorrect) referred to some surface feature (physical appearance) of the fossils. Those who chose trilobite were more likely than either of the other groups to cite a lack of familiarity either with trilobites themselves or with the other fossils in the first column where the trilobite is found. In fact, only four of the 15 people who correctly answered trilobite did so for the correct reason. The reason why these two particular fossils were chosen was described previously. However, one wonders what the result would have been if students had been presented with different columns in which a clam and snail occupied the layers where the trilobite and brachiopod exist in this drawing (assuming all other layers were adjusted accordingly to reflect possible realistic scenarios). Students would likely be familiar with both clams and snails, perhaps making the task more rather than less difficult. Alternatively, what would those same students have said if all fossil pictures were adapted to fill similar areas? Then a decision could not have been made based upon the size of the animal in the picture. Further implications of these results will be discussed more fully in chapter five. It should be noted now that correct answers may be false positives and lead us to conclude a greater understanding than may actually exist.

One caution regarding these findings is in order. Students were not directed to the particular locations of the trilobite and the two brachiopods. This was done deliberately so as not to influence responses. Some students, however, may not have seen both of the brachiopods. If that was the case it could have affected the responses of those who said the two layers were formed at the same time because they were both at the bottom. Some students did notice both brachiopods, and this created some dissonance. Cole and Megan, both university students, appeared to struggle with how to deal with the location of the two brachiopods. Megan was able to come to the correct conclusion, while Cole was not.

Cole: *I guess they would have to form at the same time.*

Interviewer: How do you know?

Cole: Because both the trilobite and the brachiopod are both located on the lowest level of two of the columns and the brachiopod is also located almost at the top of the middle column so I guess it maybe also depends on elevation of land, too.

Even though Megan correctly concluded that the trilobite layer must be older, she was still somewhat unsure of her answer.

Megan: Probably the trilobite [voice rose at end as in question]

Interviewer: How do you know?

Megan: Well, on here [brachiopod in 3rd column] I'd say maybe around the same time but the brachiopod up here [2nd column] is farther to the top so I'd guess that it was later.

4.3.2 Geoscience content knowledge and succession of geoscience and historical events

The reader may remember that sequencing a series of geological events is a task that has been used a number of times in the literature to assess how people perceive succession in geologic time. The task in this study was modelled after Trend (1998, 2000, 2001b), but unlike his task included events from ancient history and one from the Age of Exploration. A list of the events students sequenced for this card sort task is in Appendix A and also in Table 4.12. If geoscience content knowledge influences responses in the way I've suggested, we would expect to see evidence that students were using logical reasoning strategies to determine answers. The first life and first fish mentioned earlier is an example. We would expect they would do better on items for which those reasoning strategies alone could produce a correct answer than those for which geoscience content knowledge would be more essential. Furthermore, we would expect that responses would be no less accurate for geoscience events than for historical events with which an individual was not acquainted. Eighth and eleventh graders in this sample reported having studied ancient history in the past (in grades 6 and 9), as did all university participants. According to one eighth grader, however, that was practically ancient history itself.

I don't remember that one [Julius Caesar] cause that one was from 6th grade. (Ashley, 8)

Table 4.12 lists the events from the card sequencing task for all participants. Data is reported for the entire sample and is subsequently broken down by grade level in Tables 4.13-4.15. I have organised the data tables in the same manner as Trend to allow for comparison with his results in chapter five. Columns in Tables 4.12-4.15 are identical to those in his data tables. Column 2 (Rank: Consensus) refers to the rank or serial order assigned to these events by experts. Events are ordered from most recent to oldest based upon the mean relative rank assigned by participants in this study (column 3) rather than by column 2. Standard deviations are reported to gauge degree of ranking agreement among students. I have not calculated the statistical differences between standard deviations due to the small sample size. The deviations in this study will be compared to Trend's data in chapter five.

Adjacent mean differences provide a sense of how confident the group is, as a whole, in their responses. The mean rank for each event was compared with the mean rank of the event that immediately preceded it in serial order. Low adjacent mean differences suggest students are relatively unsure which of those events occurred first. For example, mean ranks for *woolly mammoths became extinct almost everywhere* and *first humans appeared on Earth* differ by only 0.10. In their explanations many students expressed confusion about which of these two events was first. In contrast, mean ranks for *Great Pyramids of Egypt were built* and *woolly mammoths became extinct almost everywhere* differ by 2.10, indicating that students, as a group, are confident that woolly mammoth extinction preceded the building of the pyramids. This does not provide any indication that students know by how much woolly mammoth extinction preceded the building of the pyramids merely that it did.

The only pair of events for which the mean relative rank of all students differed from the expert consensus was: *very first life appeared on Earth* and *first volcanoes erupted on Earth*. This was also the only pair of items switched by the university students. The volcanoes item was included specifically because it was expected it would be largely unknown to most students. Their placement of the event and their reasoning could provide insight into how students approached the task.

Geoscience or historical event	**Rank: Consensus**	**Mean rank: relative**	**Adjacent Mean difference**	**SD**
Christopher Columbus sailed to the New World	13	12.66	0.92	0.59
Julius Caesar was killed	12	11.74	0.57	0.85
1^{st} Olympic Games held in ancient Greece	11	11.17	0.91	0.92
Great Pyramids of Egypt were built	10	10.26	2.10	0.66
Woolly mammoths became extinct almost everywhere	9	8.16	0.10	1.16
1^{st} humans appeared on Earth	8	8.06	1.27	0.94
Dinosaurs became extinct	7	6.79	0.62	1.17
1^{st} fish appeared in Earth's waters	6	6.17	1.20	1.01
1^{st} volcanoes developed on Earth	4	4.97	0.57	1.48
Very 1^{st} life appeared on Earth	5	4.40	1.50	0.81
Origin/formation of Earth	3	2.90	0.97	0.51
Origin/formation of Sun	2	1.93	0.13	0.57
Big Bang	1	1.80	---	2.39

Table 4.12 Sequence of geoscience & historical events, all participants (N=35)

Geoscience or historical event	Rank: Consensus	Mean rank: relative	Adjacent Mean difference	SD
Christopher Columbus sailed to the New World	13	12.33	0.58	0.78
1st Olympic Games held in ancient Greece	11	11.75	0.25	1.06
Julius Caesar was killed	12	11.50	1.17	1.24
Great Pyramids of Egypt were built	10	10.33	2.00	0.49
1st humans appeared on Earth	8	8.33	0.87	0.79
Woolly mammoths became extinct almost everywhere	9	7.46	0.71	1.16
1st fish appeared in Earth's waters	6	6.75	0.46	1.22
Dinosaurs became extinct	7	6.29	0.62	1.74
1st volcanoes developed on Earth	4	5.67	1.59	1.97
Very 1st life appeared on Earth	5	4.08	1.20	1.08
Origin/formation of Earth	3	2.88	0.67	0.80
Origin/formation of Sun	2	2.21	0.79	0.72
Big Bang	1	1.42	---	0.90

Table 4.13 Sequence of geoscience & historical events, 8th graders (N=12)

Geoscience or historical event	Rank: Consensus	Mean rank: relative	Adjacent Mean difference	SD
Christopher Columbus sailed to the New World	13	12.73	0.82	0.47
Julius Caesar was killed	12	11.91	1.18	0.70
1st Olympic Games held in ancient Greece	11	10.73	0.55	0.65
Great Pyramids of Egypt were built	10	10.18	1.54	0.87
Woolly mammoths became extinct almost everywhere	9	8.64	0.82	0.81
1st humans appeared on Earth	8	7.82	0.91	0.87
Dinosaurs became extinct	7	6.91	1.09	0.83
1st fish appeared in Earth's waters	6	5.82	1.27	0.75
Very 1st life appeared on Earth	5	4.55	0.28	0.69
1st volcanoes developed on Earth	4	4.27	1.45	0.79
Origin/formation of Earth	3	2.82	0.00	0.40
Big Bang	1	2.82	1.00	4.14
Origin/formation of Sun	2	1.82	---	0.40

Table 4.14 Sequence of geoscience & historical events, 11th graders (N=11)

Geoscience or historical event	Rank: Consensus	Mean rank: relative	Adjacent Mean difference	SD
Christopher Columbus sailed to the New World	13	12.91	1.08	0.29
Julius Caesar was killed	12	11.83	0.83	0.39
1st Olympic Games held in ancient Greece	11	11.00	0.75	0.74
Great Pyramids of Egypt were built	10	10.25	1.83	0.62
Woolly mammoths became extinct almost everywhere	9	8.42	0.42	1.16
1st humans appeared on Earth	8	8.00	0.83	1.13
Dinosaurs became extinct	7	7.17	1.25	0.39
1st fish appeared in Earth's waters	6	5.92	1.00	0.79
1st volcanoes developed on Earth	4	4.92	0.34	1.16
Very 1st life appeared on Earth	5	4.58	1.58	0.51
Origin/formation of Earth	3	3.00	1.25	0.00
Origin/formation of Sun	2	1.75	0.50	0.45
Big Bang	1	1.25	---	0.45

Table 4.15 Sequence of geoscience & historical events, university students (N=12)

Like the university students, eighth graders switched *1st volcanoes* and *1st life.* They also switched the historical events *1st Olympic Games in ancient Greece* and *Julius Caesar was killed.* The only pair switched by eleventh graders was the *Big Bang* and *origin/formation of the Sun*. This was due to two students who placed the Big Bang very late in the sequence (one at the ninth position and one at the thirteenth) and also explains the high standard deviation for the *Big Bang* compared to the other groups. At the same time, the eleventh graders were the only group to correctly place *very first life appeared on Earth* after *first volcanoes erupted on Earth*.

Because the students in this sample were largely in agreement with the expert consensus regarding the sequence of these events, we might conclude that there are few difficulties with succession of either geoscience or historical events. However, of greater interest than whether or not students ordered events correctly are the strategies they used to arrive at their answers. As has already been demonstrated, it is possible to get the correct answer for the wrong reason. The task itself appeared to engage students' thought processes and resulted in students changing their answers. This occurred most commonly while students were explaining their sequencing. Fifteen of the 35 students changed the order of the cards during their explanations. Many individuals seemed to be employing any bit of information they knew to help them arrive at their conclusions. For example, most students correctly placed the *origin/formation of the Sun* before the *origin/formation of the Earth*. Their reasons often had to do with the fact that the the Sun provides light/heat to the Earth. While this is true, that fact does not <u>require</u> that the formation of the Sun precede the formation of the Earth. Consider these examples.

> *I picked the Sun cause you need to have the Sun to obviously to help make the Earth live so you can't have the Earth without the Sun so I picked the Sun and then the Earth. (Justin, 11)*

> *The origin/formation of the Sun, I thought that came first cause it needs to heat the Earth. (Ashley, 8)*

Rationales demonstrate significant confusion about certain events like the Big Bang, as the following students' comments show. While many of these students placed the event correctly, they displayed interesting notions about it.

The Big Bang was something, I think, that made Earth or something like that. I know cause we did that in science, but I don't know. Someone did a project. We had to do something with Earth, how it was formed or something like that. (Sofia, 8)

Jamal (8): I picked the Big Bang because I think that's when the Earth had been formed.

Interviewer: What do you know about the Big Bang?

Jamal: That it was a large asteroid and it hit another one I think, and it became one huge planet.

Elizabeth (univ): For the first two, the origin of the Sun and the Big Bang, I'm not sure if the Big Bang actually created the Sun or not or if it existed before but I guess it did exist before. It was just like a guess.

Interviewer: You guess that the Sun existed before the Big Bang.

Elizabeth: Yeah, I know I read about the Big Bang but I can't remember if it was what created everything in the universe or if the Sun already existed before that.

Students not only displayed confusion about the timing of geoscience events but also more "recent" historical events. Hannah confused the timing of Julius Caesar's death with the Shakespearean play of the same name.

Hannah (11): Then I remember I read the play [Julius Caesar]. I remember a lot about that. I'm not completely sure on the year but I feel like it would have *been more modern than, oh well, duh, it had to come after Christopher Columbus because of Shakespeare. Shakespeare was obviously after Christopher Columbus's time.*

Several students, such as Michael (11), mentioned movies when justifying their responses, specifically the film, *10,000 BC* which unfortunately fosters the alternative conception that woolly mammoths were used to build the pyramids of Egypt.

Michael: Then I watch the Discovery Channel too much so I think woolly mammoths died after the dinosaurs. Then the humans [switched humans & woolly mammoths to place 1st humans before woolly mammoths].

Interviewer: Why did you decide to switch the humans and woolly mammoths?

Michael: 10,000 BC. Then the pyramids, no! [switched pyramids & mammoths to place pyramids before woolly mammoths]

Interviewer: And why are you switching those?

Michael: The movie [10,000 BC]

Interviewer: What's in the movie that would make you think that?

Michael: They would capture the head of the pack of woolly mammoths and the others would just follow and they used them to carry the blocks to build the pyramids, but I don't know if it's right cause that was a crazy movie.

When explaining their sequencing, a number of students simply recited their cards in order without explanation and had to be prompted to provide reasons for their choices. It is difficult to know if they were simply reluctant to spontaneously express their reasons or if their placements were not well thought out and the reasons they gave in response to prompting were generated only after that prompting. Often students appeared to combine specific knowledge of events with logical reasoning to order the cards.

> *Then life obviously would have to come after that because no living thing would be able to happen if there wasn't the 1st. Then fish because it's an animal and I know that animals have been around a lot longer than humans have and I know that they date back to dinosaurs and stuff like that. (Hannah, 11)*

> *That [first life] would be before the fish. Then the fish because obviously if the first life appeared that wouldn't be as complex as a fish. (Anthony, univ)*

> *I'm not sure if mammoths became extinct before humans came or after. It could probably be after too cause they aren't as dangerous as dinosaurs so it could be that they were living some when humans appeared. (Claire, univ)*

Claire's response espouses a view that all dinosaurs were dangerous in contrast to the more docile woolly mammoths.

While participants were often able to successfully use some combination of specific knowledge and logical reasoning strategies to correctly order events, sometimes the combination resulted in rather fanciful responses such as Danielle's (univ),

> *The first volcanoes developed on the Earth. Well the Earth needed a little time to start colliding with the plates which would form volcanoes. Then I would*

imagine that all of the volcanoes going off would kill the dinosaurs and once it killed off the dinosaurs it would also kill some of the woolly mammoths. But at the same time once everything had kind of been destroyed by the fires from the volcano it would get cold and end up killing the rest of the woolly mammoths which would make it extinct. Then people are still around. I'm not sure how they survived the volcano unless you take the Flood [Noachic Flood in the Hebrew Bible]. *So anyway you have people.*

Twenty-six participants specifically said they were guessing or weren't sure about some of their answers. They were more likely to say they were guessing on the later ancient history items than with the earlier geoscience events. Confusion about the geoscience events centred on *1st volcanoes, extinction of dinosaurs, extinction of woolly mammoths, and appearance of 1st humans.* Based upon mean relative rankings and standard deviations, students were able to sequence geoscience events as well as they sequenced historical ones. The only standard deviation that is a bit larger than the others is the *Big Bang.* The reason for this has already been mentioned. Explanations suggest that an item such as *first volcanoes erupted on Earth* were difficult because participants were unable to use logical reasoning strategies to place it.

Student explanations for their sequencing revealed a number of alternative conceptions, both scientific and historical. Some have already been mentioned. A few others are printed below.

It's probably just assumed that fish came before dinosaurs because one fish grew to a big dinosaur. (Alyssa, 8)

Then 1st volcanoes developed on Earth that's like a geographical thing, a geographical feature. I guess after Pangaea there came volcanoes from the plates shifting and maybe on the other side of the earth where Pangaea wasn't cause of the plates, there were volcanoes maybe on that side, so that's why I put that there. Then the 1st life, I think you need that for pretty much everything else to happen. Then dinosaurs, I put that because I think they were the earliest forms of, not the earliest life but maybe the earliest form of advanced life like up from the single or double celled organisms or like smaller things. I think the dinosaurs kinda started all the animals that we know and that's also why I put the fish there was because maybe the fish developed after the dinosaurs became extinct. There was DNA left from the amphibious dinosaurs and that's why I put the fish there. (Connor, 8)

I think the Olympic Games were held really early on AD and Julius Caesar was killed in mid-400 definitely, I think. I don't know, I just remember the Roman Empire being around 300, 400. (Ryan, 11)

What emerges is a mishmash of student ideas about events that they tried to make sense of to complete the task. Frequently students took less than accurate domain or topic knowledge, coupled it with reasoning strategies, and placed events in the correct order. Events that were closer in time, like the ancient history events were more problematic because specific subject matter knowledge was required to distinguish among them. Many of the earlier geo-events could be sequenced on the basis of logical reasoning strategies, such as that displayed by Malik (11).

Then I chose the formation of Earth next because there's no place for them [other cards] *to be 'cept on Earth.*

Following the card sort, students were asked to group cards into categories of their choosing and to provide a name for their groups, if possible. This is similar to the methodology employed by studies investigating student conceptions of size (Jones et al., 2009; Jones et al., 2008; Tretter et al., 2006; Tretter et al., 2006). Number of groups ranged from four to ten with a mean of 6.97. Not all students were able to provide names for all groups, and one eighth grader was unable to provide a name for any group. Those who did name the groups rarely mentioned age as a descriptor, suggesting that time was a minimal factor in their thinking about the task. There was evidence of the compressed logarithmic scale described in section 2.3.1. Finer temporal distinctions were made between events that were more recent (ancient history events were usually in a different pile than Columbus) while the first three events were often grouped together. Students were then prompted to put an age name on as many groups as possible. Again, many students had great difficulty with the task. Seven students were unable to attach an age name to any group, while 20 students did so for at least one group. The remaining eight provided age names for all groups. Those age names varied widely.

The oldest age name given for the oldest group was hundreds of billions of years ago, well beyond the accepted date for the Big Bang, while the youngest age name for the oldest group was 2,500 BC. Nine students said they were unable to provide an age name for the oldest group. Of the 26 who provided a name, 12 of those said something in the billions, eight said something in the millions, and six said something that involved BC [No one used BCE]. Only fourteen individuals gave an age name for the pile containing first humans. Those dates ranged from 5-65 million years ago to 2,000 BC. Ages suggested for the pile containing Columbus's voyage to the New World were more accurate, but there was still wide variation with one student saying the date was 1942. Many U.S. students are taught a rhyme to remember the date of Columbus' voyage that begins, "In 1492, Columbus sailed the ocean blue." Several students quoted that rhyme when trying to determine a date for his voyage to the New World. It is quite possible that the student (Emma, 8) who said 1942 simply mixed up the digits in the date. However, after she said 1942, it was pointed out to her that 1942 was around the time of World War II. She was asked if that was the date she meant for Columbus' voyage. Emma immediately replied that it was. On that basis, I think it is reasonable to conclude that she does not have a good sense of where that event sits in time. It is not surprising that students have difficulty placing an absolute age on geoscience events no matter how the task is structured. This study suggests that it is not only events in deep time that students cannot place, but that they have difficulty placing historical events as well.

4.3.3 Geoscience content knowledge and duration

One unique contribution of this thesis is the investigation of how students understand duration in deep time. Two tasks were employed to answer this question. By their nature, geoscience events are often unfamiliar to students, so it would not be surprising if they had poor understanding of the duration of those events. This would easily be attributable to a lack of knowledge. However, if they demonstrated an equally poor understanding of events that occur in conventional time with which they are unfamiliar; this would suggest that the lack of familiarity with the event itself is a critical factor and not merely the length of the duration.

4.3.3.1 Duration of events questionnaire

The items in the duration of events questionnaire can be found in Appendix A. The questionnaire was used in two ways. The first was to compare accuracy of responses for items that occur within conventional time with those that occur outside its bounds. It was hypothesized that items of short duration (seconds, minutes, and days) that were unknown to students would be just as difficult for them as geologic events of longer duration. If students are similarly inaccurate in their judgements of durations for unfamiliar events that require very different durations, this suggests that the most salient factor in their judgements is not the length of time itself but rather their familiarity with the event, i.e., their subject matter knowledge. Of course, it is theoretically possible to get some answers correct without specific subject matter knowledge but by relying on logical reasoning strategies. That is potentially true for items of both short and long duration. There is no expectation that logical reasoning would be of greater help with items of longer as opposed to shorter durations or vice versa. The second was to triangulate responses between this item and those on Timeline 3 which will be discussed in the next section. These tasks had five items in common. If a student indicated a time period for the duration of an event here that was different from the duration that was indicated in the timeline that would raise questions about the reliability and validity of responses.

As mentioned in chapter three this item was modelled after one used by others to assess students' conceptions of size (Jones et al., 2009; Jones et al., 2008;

Tretter et al., 2006; Tretter et al., 2006). There are several cautions in interpretation, two of which are common to their work with size and one of which is more specific to time. First, it is possible to check the correct unit box for the duration of an event yet still hold a fairly inaccurate conception regarding the event's duration. If a person checked the *years* box for the amount of time necessary for the Voyager probe to reach Jupiter, he could be reasoning that the probe takes up to 99 years to reach Jupiter when it actually takes two years. His personal conception of the event's duration may still be quite inaccurate. Checking the correct box only indicates the extent to which an individual perceives the correct unit of time with which to express the duration of a particular event.

Some events on the questionnaire have the potential for greater variability in the estimation of their durations than others. An event that takes millions of years to occur could only be underestimated, not overestimated on this task since *millions of years* was the greatest time period possible. An event that takes seconds could only be overestimated, not underestimated. In contrast, an event that takes years could be both under- and overestimated. Both of these problems are shared with the size studies mentioned above.

Estimating durations raises an additional issue of interpretation that is not a factor when dealing with size. Since students in the scale studies were asked to indicate the size of objects using metric measurements, the researchers were able to determine by how many powers of ten students had over- or underestimated the size of an object. The difficulty with units of time is that they do not use the same base (60 minutes in one hour, 24 hours in one day, etc.). Thus, it is not reasonable to quantify over- or underestimation in terms of the number of scale factors by which the estimation was off. Converting time units to powers of ten makes little sense since we don't think of units of time in that way. One doesn't conceive of one second as being 0.0167 minute. These three issues indicate that data from this item must be interpreted cautiously. Nonetheless, it is a useful item for the two reasons mentioned above.

A tally was created for each item indicating the number of students who estimated the duration of an event with the accurate unit of time as well as those who over- or underestimated and by how many units of time they did so. The percentage of students who indicated the correct unit was determined for each event. Table 4.16 shows items on the questionnaire whose durations were estimated most accurately and those that were estimated least accurately. The responses of three students were excluded from this analysis because they indicated a different duration for an item on the questionnaire than they did on Timeline 3. Those discrepancies will be discussed in the next section. Excluding those three students made no difference in ranking of these events.

Item	Percent estimating duration accurately	Item	Percent estimating duration accurately
Eat dinner	100%	Red blood cell to make one trip throughout body	14%
Fly from New York to Los Angeles	94%	Colorado River to carve the Grand Canyon	17%
Amount of time ground shakes during an earthquake (seconds & minutes were both counted correct)	91%	Build the Great Wall of China	26%
Pumpkin grown from seed to be ripe	89%	Light to travel from the Sun to the Earth	29%
Drive from one side of Pennsylvania to the other	83%	Appalachian Mountains to form	29%
Hair on head to grow ½"	83%	Sedimentary rock to form	34%

Table 4.16 Most and least accurate estimates on duration of events questionnaire (N=35)

All the events that were estimated most accurately occur well within the bounds of conventional time. In fact, none takes longer than days to occur. One of those is a geological event (ground shaking during an earthquake). While they may not all have been experienced by every individual in the study (e.g., earthquake), these events are all common to human experience more broadly. In contrast, there is great variety in the durations of the events that were least accurately estimated. They range from less than one minute for a red blood cell to travel through the body to millions of years for the Appalachian Mountains to form. Three of the events in this group are geological and one is astronomical. The latter (amount of time for light to travel from the Sun to the Earth) demonstrated the greatest variability of any item in the questionnaire. One person checked seconds, and three others said it takes millions of years for the Sun's light to reach the Earth. These results must be interpreted carefully, but they seem to indicate that specific knowledge of an event is at least as critical in estimating its duration as the length of the duration. Events that occur within conventional time are just as problematic for students as those occurring within deep time if both are unfamiliar.

4.3.3.2 Timeline 3: Geoscience content knowledge and duration with large numbers

Section 4.2 reports on the results of three numeric timelines. The ability to correctly place times on those number lines involved knowledge of the numbers used and the proportional relationships among them. No geoscience knowledge was required. Timeline 3 was different in that students were instructed to place events on a timeline based upon how long they take to happen in proportion to each other. Based upon the model of the"three-legged stool," we would expect that students would more accurately place the durations of events with which they were familiar and less accurately place those with which they were unfamiliar. Further, events in conventional and deep time that are unfamiliar should be placed just as inaccurately. Surface features are likely to be used to place them. Events are reprinted in Table 4.17 for the reader's convenience.

A.	The Earth spinning around once
B.	How long most coral reefs have been growing
C.	The break-up of the supercontinent Pangaea
D.	The Earth going around the Sun once
E.	The Moon going around the Earth once
F.	The carving of the Grand Canyon by the Colorado River
G.	The amount of time the ground shakes during an earthquake

Table 4.17 Stimulus items for Timeline 3

Timeline 3 was not scored for correctness. It was not expected that anyone would be able to order all these events in a proportional linear scale. It seemed likely that many students would be unsure about the durations of several of the events, particularly coral reefs, the Grand Canyon, and the break-up of Pangaea. Instead, attention was paid to the reasons students gave for the placement of letters and any mention of duration in their explanations. It was expected that students would mix up the order of two unknown events. What was considered important was that timelines reflected student explanations and that there was a clear, proportional distinction between the events that take one year or less and those that take much longer. For example, if a student incorrectly said the Colorado River took thousands of years to carve the Grand Canyon and placed F on the timeline in a place that would make sense based upon that duration, that timeline would indicate an understanding of large numbers but a lack of knowledge about the event itself.

Participants displayed many alternative conceptions about the durations of shorter events as well as longer ones. Some said the Moon orbits Earth in one day, equating it with a day/night cycle. Others said the Moon's orbit requires one year.

One person thought the Earth spins around once in one hour. Again, it was not surprising that many were unclear about the durations of B [coral reefs], C [Pangaea], and F [Grand Canyon], but some estimates for their durations were especially inaccurate. A few indicated the break-up of Pangaea took a few years or months, and one person stated the Colorado River carved the Grand Canyon in a week.

The rationales participants gave for their estimated durations are even more enlightening. Several people equated size with duration. This was evident in what they said about the Grand Canyon. Two said that the great depth of the canyon required a long duration. When Peter (univ) compared the Grand Canyon to coral reefs he said that the Grand Canyon would take longer because it's larger.

Others focused on the characteristics of the materials. When explaining why the Grand Canyon took a long time, Hannah (11) said,

> *You know it's gonna take awhile cause land isn't soft like, it's not in all spots, at least not here so there's no way it could have just broken up easily or in a short amount of time. It would take a long time to get to the way it is.*

The comparison between the durations of coral reefs and the Grand Canyon, in particular, generated a lot of interesting responses related to characteristics of the materials. Two students said the growth of coral reefs would take less time than the Grand Canyon because corals are living organisms while the Grand Canyon has to do with erosion [one used the term weathering instead]. The presumption is that biological processes require shorter durations than inorganic ones. A university student who thought coral reefs take less time to grow than the Grand Canyon attributed her answer to her understanding that "stuff in the ocean is older than stuff on the land," a view not solidly eradicated from the scientific community until the advent of deep sea drilling.

Vincent's (univ) responses suggest some underlying distrust of the validity of uniformitarianism, a cornerstone of modern geology. When placing Pangaea he said it, "takes a long time, but not as long as they think." In talking about the Grand Canyon he said, "At today's rate it takes a long time." His placements may indicate the presence of metaphysical confounding factors that affect the stability of the deep time "stool" that were described in section 2.8.

Most striking of all, many students appeared to be trying to latch onto any bit of information they thought they knew that could help them answer the questions. Just as was the case for the card sort sequencing task, this resulted in very interesting responses. In talking about the Grand Canyon, Chris (8) said,

> *The Grand Canyon, definitely a long time because I remember reading or hearing somewhere that maybe in about 50 years that Niagara Falls is gonna be flat, so I heard something like that so the Grand Canyon with that deep it's gonna be really a lot of time.*

Jamal (8) described the break-up of Pangaea as "moving a big island and pushing it into a different place." Danielle attempted to determine whether coral reefs took longer to form than the Grand Canyon. She ultimately decided that was the case and offered this rationale.

I would guess longer than the Grand Canyon just because I don't honestly know where the source of the Grand Canyon is so I imagine it would depend on how it started out, like if it started out as a lake or something and then the river started to flow out from it or if it's an underground thing.

Once again, we see great confusion regarding these events. This was true not only for geoscience events but also for events of shorter duration.

4.3.3.3 Comparison of Timeline 3 with duration of events questionnaire

Durations indicated on the questionnaire were compared to those on Timeline 3 for the items that were common to both in order to triangulate responses and monitor consistency of responses. Those common items were:

- the amount of time the ground shakes during an earthquake
- the Moon going around the Earth once
- how long most coral reefs have been growing
- the carving of the Grand Canyon by the Colorado River
- the break-up of the supercontinent Pangaea

Table 4.18 lists the number of students who mentioned a specific duration for these events as part of their explanation for Timeline 3. The duration could have either been mentioned spontaneously or in response to a question by the interviewer. In order to be counted as having mentioned a specific duration, the student needed to mention a unit. "A really, really long time" was not counted. "Years" or "hundreds of years" was counted. If all students had mentioned durations in their explanations for those five items in the timeline a total of 175 durations would have been cited. What is striking is the fact that students named few durations for events even after prompting by the interviewer. There were only 58 instances when a student mentioned a specific duration for one of these items. In contrast, no one left any item blank on the questionnaire, which was a forced choice format.

Number of durations suggested	**Number of students**
0	9
1	9
2	8
3	5
4	2
5	2

Table 4.18 Number of students who suggested durations for five events as part of their explanation for Timeline 3 (N=35)

Because few students mentioned durations for events on Timeline 3, only limited correlation with their responses on the questionnaire was possible. Of the 58 durations mentioned for Timeline 3, there were eight instances in which a person named different durations for an event on the timeline and the questionnaire. Two of those could be attributed to the forced choice format of the questionnaire. Justin said that most coral reefs have been growing for one billion years while constructing Timeline 3. Anthony said that Pangaea took a billion years to break up. Both checked millions of years for those items on the questionnaire (the largest choice possible). Rather than signalling an inconsistency in their responses, it is more likely a reflection of the fact that a response they might have chosen was not available to them.

Three other students had discrepancies between the durations they mentioned on Timeline 3 and the questionnaire. Two of the students were eighth graders. Matt said it took the Colorado River 500 thousand years to carve the Grand Canyon while explaining Timeline 3. He checked the box *hundreds of years* on the questionnaire for the same event. He also said most coral reefs have been growing for ten million years but checked the box *thousands of years* on the questionnaire. Jamal indicated that coral reefs have been growing for either 100 or one million years, but checked *thousands of years.* He said the Moon orbits the Earth in one day but checked the box *years.* Elizabeth, a university student, said coral reefs have been growing for thousands of years, but she checked the box *millions of years.* She said the ground shakes for minutes during an earthquake but checked the box *seconds.* The responses of those three students were not counted when determining the accuracy of responses for the questionnaire (see Table 4.16) since their answers were not consistent across tasks. Other than these exceptions, student responses regarding durations were consistent across the two tasks. This does not mean individuals were confident of their responses or that they weren't guessing. In fact, many students said they <u>were</u> guessing. It does suggest that their guesses were not random and that their guesses were consistent across tasks.

4.3.4 How does geoscience content knowledge impact how these students understand deep time?

In chapter two, I noted two ways in which a student's geoscience content knowledge could be a factor in deep time. The extent of a person's declarative and procedural knowledge will determine how new information is perceived. In the absence of solid geoscience content knowledge a student's judgments of both succession and duration will likely be made on the basis of surface features such as physical appearance. The final research question asked whether students in this sample relied on surface features to make judgments about duration in deep time instead of citing geoscience ideas for their responses. That was, in fact, the case for the items in this study. Even students who gave correct answers frequently cited surface features as justification for responses, making their correct answers coincidental. Paying attention to those same surface features could lead to an incorrect answer in another instance. In fact, students who answer correctly and those who answer incorrectly sometimes do so for the exact same reason. If we are not careful, we may assume students who respond correctly possess a better understanding than is actually the case.

In both succession and duration tasks, students appeared to have few subject specific referents around which to base their answers. Hence, their knowledge was often fragmentary and confused at best. No one displayed evidence of accurate conceptions in all areas, which was not at all surprising since there were no geoscience experts in the sample. Participants often recognised a term but displayed little understanding of the event associated with the term. The Big Bang was associated with the formation of Earth as well as an asteroid impact. When asked to provide specific ages or durations for events, the numbers provided varied widely across participants. Events that occur within human timescales appeared to be no easier for students to judge than those that occur in geologic time <u>if</u> they were unfamiliar with both.

As expected, duration appears to be a far more difficult concept for students to understand than succession. There was some distinction between physical processes and biological ones, with physical processes being deemed to require longer durations. Size was often equated with duration. Hence, the carving of the Grand Canyon was judged to require a longer amount of time than the growth of a coral reef since the former is larger. With a few exceptions, students in this sample were consistent in their duration judgments across tasks. This cannot be construed to indicate they were confident of their responses. It does, however, suggest that answers were dependable.

4.4 Relationships among the three "legs" of the "stool"

To what extent might success in one area be correlated with success in another? That is, do students who do well in one area of the interviews tend to do well in other areas? The next few sections explore relationships among several of the tasks across the interviews. Each of the comparison tasks was chosen because it discriminated participants in some way. Individuals performed so well on the succession tasks in conventional time that no real groupings emerged. Duration tasks in conventional time did discriminate. Thus, duration items for short time periods (animations) were compared with those for longer periods (numeric timelines). Durations involving numbers ranging from small to very large (Timeline 4) were compared with durations of the same time periods that added subject matter knowledge as a variable.

4.4.1 Comparison of duration animations with the numeric timelines

Data from the three numeric timelines was compared to results for the duration animations at the individual level to see if students who had difficulty with conventional time had similar problems with large numbers. A scatter plot was created to show the relationship between the two tasks. In order to create the plot, numeric values were assigned to the timeline groupings: three for the sufficient category, two for the insufficient, and one for the poor category. This was done solely to allow for the creation of the graph. It does not imply that the categories are interval. In fact, they are not; they are ordinal. Variability exists in the ways a student could earn four points on the animations. This graph does not distinguish between those ways. (Someone who answered both questions after A1 and A3 correctly may be different from someone who answered two correctly after A2 and one after A2 and A3.) Thus, this data can only suggest that there may or may not a relationship between performance on the animations and the numeric timelines. Data points on Figure 4.21 are different sizes to denote the fact that not all points on the graph represent the same number of students. The dot at the ordered pair (1, 3) is very small. That is because only one person answered one question correctly after the animations and was ranked in the *sufficient* category on the timelines. In contrast, the dot at the ordered pair (4, 3) is quite large. Nine people answered four questions correctly after the animations and were ranked in the *sufficient* category on the timelines.

There does not appear to be any relationship between how many questions a participant answered correctly after the animations and whether or not their understanding of large numbers is sufficient to deal with deep time [$r(33) = 0.24$, *ns*]. Performance on one task would not be a good predictor of performance on the other.

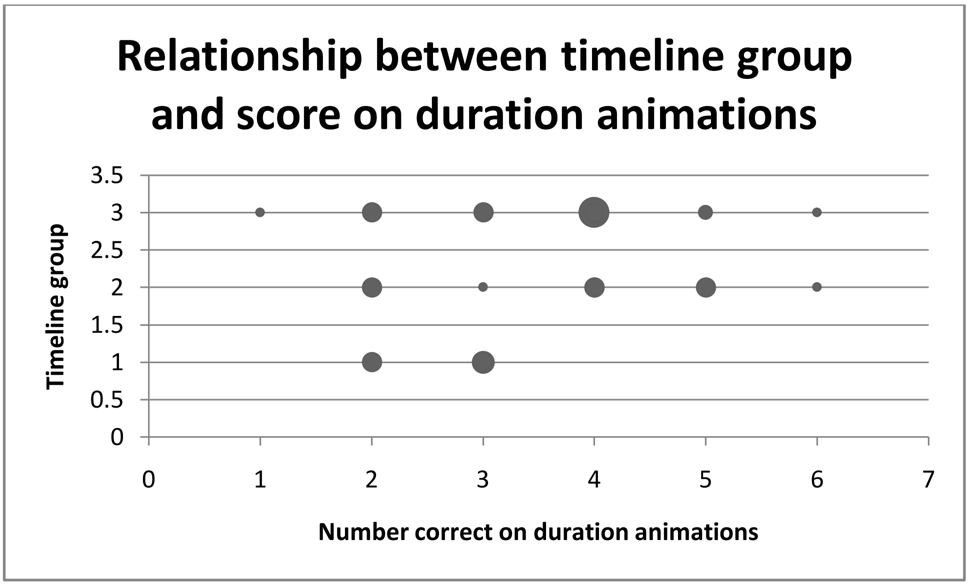


Figure 4.21 Comparison of performance on animation questions with performance on Timelines 1, 2, & 4 (N=35)

4.4.2 Comparison of Timeline 3 with Timeline 4

In order to compare student understanding of large numbers with subject matter knowledge, Timelines 3 and 4 were compared to determine what, if any, differences existed between them. In both, students had to place the durations of seven items on a linear, proportional scale. TL3 required knowledge of the duration of specific events as well as the proportional relationships among those time periods. TL4 only required knowledge of mathematical relationships among powers of ten or units of time. The reader will recall that numbers in TL4 were chosen because they are the durations of the events in TL3 rounded to the nearest power of ten or unit of time.

Students who possess insufficient knowledge of large numbers to deal with deep time or a poor understanding of smaller numbers had great difficulty with TL4.

In fact, it would be practically impossible to determine if a student ended up in the poor category or the insufficient for the numeric timelines based solely by inspecting their TL4. There is really no difference in the timelines drawn by students in those two groups. The majority of these individuals spaced events for TL4 fairly evenly across the line. These two groups constructed a TL3 that was practically identical to what they had done for TL4. A comparison of Vanessa's (11) TL3, printed here, with Jenna's (8) TL4 which was reproduced in Figure 4.4 (p. 245), shows that for these students there was no difference between TL4 (numbers alone) and TL3 (events).

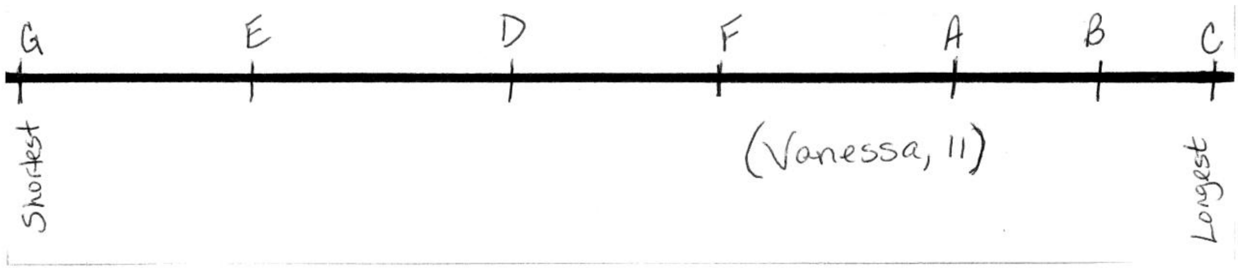


Figure 4.22 Vanessa's Timeline 3

It is not possible to say much more about those first two groups beyond the fact that they have trouble with large numbers and with relevant geoscience content knowledge. It is unclear, at present, if one of the two factors is more problematic than the other. Both tasks may simply have been too difficult for them. Even those who had successfully placed four numbers for TL1 were unable to place even the smallest durations proportionally on either TL3 or TL4. The fact that some were not sure about the duration of shorter events (Alyssa said the Earth revolves around the Sun in a day) makes it difficult to argue that geologic events that occur over long timescales are necessarily more difficult to comprehend than those of shorter duration.

Students with sufficient knowledge of large numbers to deal with deep time who had difficulty with TL3 lend support to the idea that while a knowledge of large numbers is crucial, if a student does not also possess relevant geoscience content knowledge, it will be difficult to develop meaningful understanding of the time required for processes in deep time. As we saw in section 2.2.4, these students differed from the others in the sample in their problem solving approach to the timelines. This group tended to use the same strategy for TL3 as they had for the numeric timelines. Michael (11) is an example of a student who employed a benchmarking or reference point strategy on both TL3 and TL4. His timeline TL4 is reproduced in section 4.2.4 (Figure 4.14, p. 258).

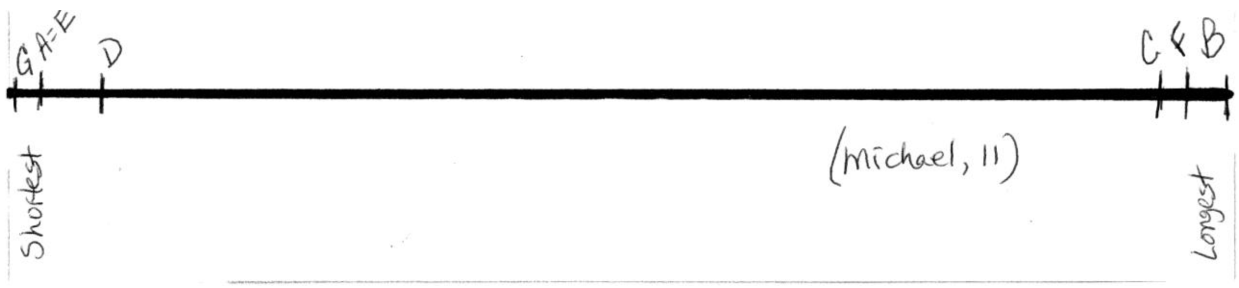


Figure 4.23 Michael's Timeline 3

Prior to placing the letters on the timeline, Michael made a list of the events, and put a box around D, the Earth going around the Sun once. He was trying to benchmark each of the events against a known duration for D as this interchange demonstrates,

> Interviewer: I have a question that has to do with what you wrote here. Why did you put a box around D?

Michael: D? Because D takes a year and all these are really short and all these are really long so I put that right in the middle. It's important cause it's so short compared to these.

Ryan (11) is an example of someone who assigned temporal markers to each of the events, a strategy that was common amongst these students. His TL4 can be found in section 4.2.4 (Figure 4.17, p. 260). While explaining TL3 he indicated that he was unsure of the durations of some of the events. Like Michael, he combined what he knew about large numbers with his ideas about the durations of these events to complete TL3.

Ryan: I took the longest one which was the break-up of Pangaea and brought that over to the end. I just took a pretty random guess about how many years ago I thought that was. I just went with like about 100 million. Then I went down to coral reefs and I had to guess at how long they've been growing cause I don't know anything about coral reefs and I just guessed about maybe 60 million years. Then I went down to Grand Canyon cause I know for a fact that took awhile so I'd say at least, well I probably could have brought it back a little further, but at least through the 100 thousands. [moved F farther to left]

Interviewer: So why would it make sense to move it to the left if it takes thousands of years?

Ryan: Cause shortest we're talking thousands of years to longest which would be millions of years.

Interviewer: You squeezed the last ones very close together

Ryan: Yeah, cause relative to the timeline that's like the blink of an eye

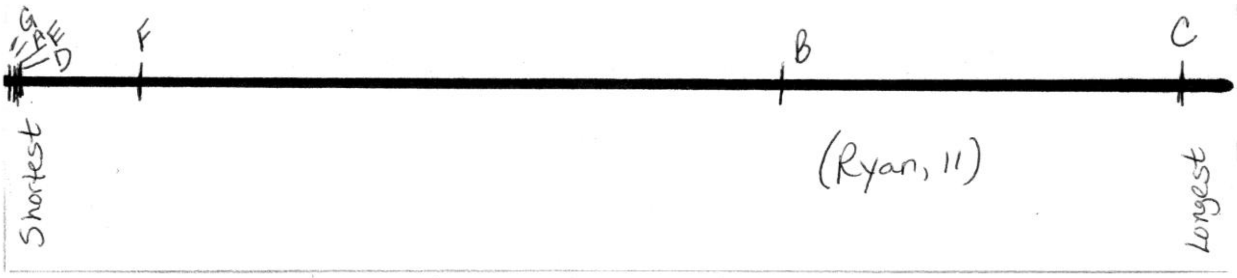


Figure 4.24 Ryan's Timeline 3

Ryan's TL3 is somewhat unique amongst this group as many of these individuals did not discriminate durations for the coral reef, Grand Canyon, or break-up of Pangaea. There were gaps in the knowledge base in this group. Four of these students said the Moon orbits the Earth in one day while explaining TL3. That alternate conception doesn't show up on this scale but would result in an inaccurate timeline on a smaller scale.

Just because students in this group used a numeric strategy did not guarantee they would produce a proportional TL3. Sarah (univ) mentioned numbers for the durations of the events for TL3, but did not use the proportional relationships among those numbers to the same extent she did with her other timelines.

Sarah: Then it takes a year for the Earth to go around the Sun so that's next. I think it took a really long time for the coral reefs to have been growing so we'll say a couple million years for that and a couple million years for the Grand Canyon so they're close to each other anyway. Then the break-up of the supercontinent Pangaea is definitely the longest because it took millions, it took a long time so that's the farthest away.

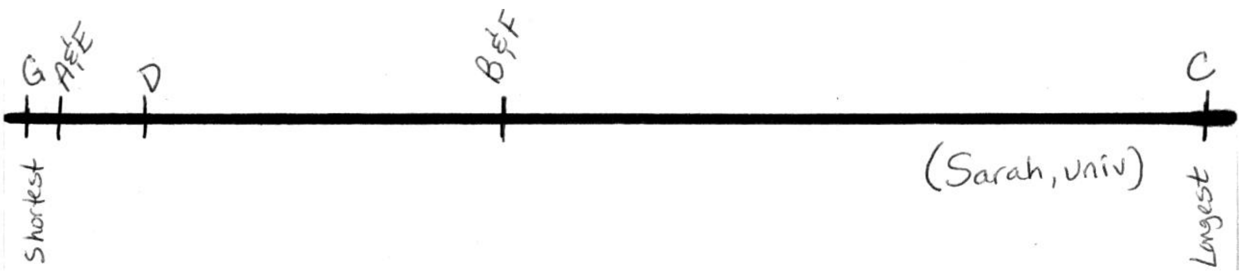


Figure 4.25 Sarah's Timeline 3

Contrast TL3 with TL4. The year is placed very differently on the two timelines. The shortest times including one year are bunched very closely to the left side of the timeline for TL4. On TL3 she creates much clearer distinctions among the smallest durations than she does for the fourth timeline. This is true even though she says A and E are both one day and D is one year. Sarah maintains better proportionality between ten thousand years and ten million years and between ten million years and 100 million years on TL4. On TL3 she says the Grand Canyon took a couple million years but places it closer to Pangaea than she placed 10 million years to 100 million years on TL4. At least in her case, the addition of the events themselves appears to have made the task more difficult.

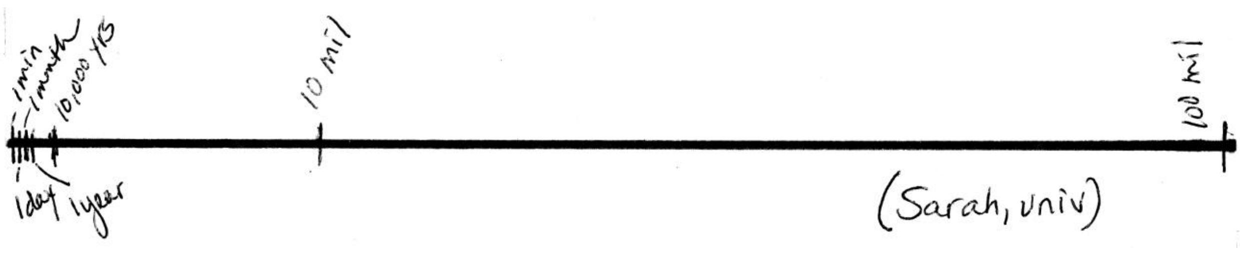


Figure 4.26 Sarah's Timeline 4

David has been mentioned before as someone whose verbal explanations suggested a more sophisticated understanding than what the timeline alone indicated. As with his numeric timelines, TL3 was somewhat problematic. In this case, he didn't even use the entire timeline and proportionality across the line was inconsistent. Even though he didn't get factual information entirely correct, he showed a good understanding of how to view shorter events in light of the scale necessary for this timeline. The timeline alone without his explanation would not lead one to conclude his understanding was very good at all.

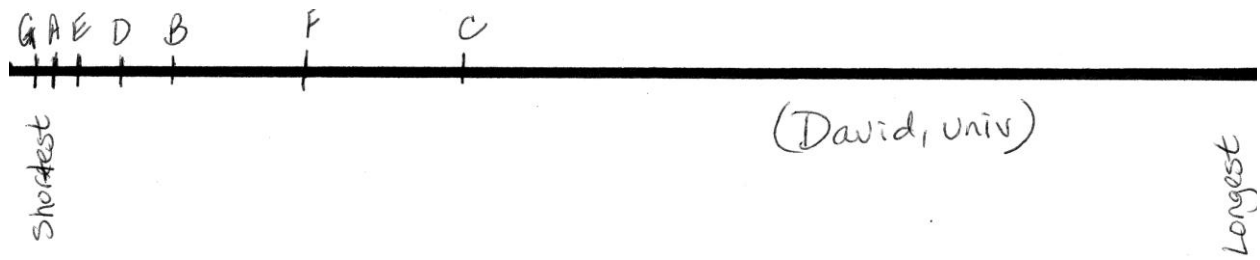


Figure 4.27 David's Timeline 3

David: I put G 1st which is the amount of time the ground shakes during an earthquake which is relatively short. Then I put the Earth spinning around once which is basically a day. Then I put the Moon going around the Earth which is about 23 ½ days. Then I put the Earth going around the Sun once which is one year. Then I put how long most coral reefs have been growing. Now I don't know how long it's been growing but I figure it's more than a couple years considering the fact that it takes a plant like a tree, like 10 years to grow. I've heard some things about how they've been growing for a long time I just, I couldn't put a number on it. Then I put the carving of the Grand Canyon by the Colorado River and then the break-up of the supercontinent Pangaea. I don't *know why I put those there like that cause I don't know how long it took either of them to do that but I'm gonna guess that the Grand Canyon took a little bit of a shorter time than the break-up of Pangaea.*

Interviewer: You said the Moon going around the Earth once was around 23 days.

David: Something like that.

Interviewer: And you said the Earth going around the Sun once was a year. You put them very close together on this timeline. How come?

David: Because, umm, well, why I put them close together. Relative to how long it takes for Pangaea to break up, the plates are moving 2 cm a year, how long does that take to move, a long time basically. Relative to how long it took for the carving of the Grand Canyon or the break-up of Pangaea, it's like a day. It's like if it was my life, if C, the break-up of the supercontinent was equivalent to my life, it would be like 7 seconds. It's very short. That's probably wrong, but still.

Many individuals in this group said TL3 was more difficult than the other timeline tasks because they didn't know how long the events took. In other words, they found the task difficult due to their lack of geoscience content knowledge. Ryan (11) described the difference between TL3 and TL4 after he finished the fourth line.

Interviewer: Let me ask you a question. Of the last two timelines you did, did one of those seem easier or where they about the same?

Ryan: Yes, definitely numbers, just straight numbers.

Interviewer: Why do you think that is easier?

Ryan: Just for the fact that I don't know the exact dates for the other ones. Had I known it might have been a little easier and plus all of those were factors of 10, well except for 1 minute and stuff, but I can just put it on a timeline thinking this is 10 times more than that, 10 times more than that as opposed to just a word I don't really know when exactly it occurred plus I don't think anyone knows for sure exactly when those events occurred so it would just be what I thought.

4.5 Summary of the results: Is there a "typical" 8th, 11th grader, or university student?

Students' performance varied greatly across tasks and within the same age group throughout the interview. There may be some slight age effect to the findings as university students as a group did somewhat better across all tasks than eighth graders. There were, however, exceptions and the trend is not large if it does exist. I will first review results across the entire sample and then briefly discuss specific age groups.

Only one individual had any real difficulty with succession tasks in conventional time. Results for duration tasks were mixed, with some students performing very well and others not well at all. Fifty-six percent of the students in this study had enough difficulty with large numbers to suggest they would have problems understanding deep time on that basis alone.

Students who were better able to use logical reasoning to determine answers appeared to do better across the tasks. Those who were able to weigh rate against size in the animations were more successful than those who focused on only one factor. Many individuals used logical reasoning strategies to complete the card sorting task. It was not unusual for someone to express greater confidence about his placement of earlier geoscience events than later historical ones. Similarly, students who completed timelines by starting with the largest number and working backwards completed more accurate timelines than those who started with the smallest numbers and tried to work their way up to the larger ones in the set.

Sometimes being asked to deal with the geologic context as well as something else was problematic for students. As described above, Sarah mentioned durations for Timeline 3, but didn't use the proportional relationships among the numbers in the way she had on the other timelines. This was also seen in the application of duration to the stratigraphic sequence discussed in section 4.1.3.6. Overall, the question appeared to confuse a number of participants. Even students who correctly concluded that the drawing itself didn't contain enough information to judge the length of the deposition period often gave multiple answers before settling on the correct one.

It is extremely difficult on the basis of this limited exploratory study to describe the responses of a "typical" student within any of the age groups in the sample. The study was not designed to answer that question. I offer only tentative suggestions about ways in which one age group in the sample might differ from the others. Overall, there are minimal differences across age groups consistent with previous literature on deep time conceptions (section 2.6). There are a few exceptions to the previous statement. All university students displayed a solid understanding of time periods up to 100 years. About 2/3 of them understood large numbers well enough to be able to comprehend the time periods involved for deep time. Eighth graders were over-represented among those who posses either *insufficient* or *poor* understanding of large numbers to deal with deep time. Eleventh graders were about evenly split between those whose understanding of large numbers would be sufficient to grasp deep time and those for whom that was not the case.

Succession errors in conventional time occurred with greater frequency among 8th graders than either 11th graders or university students. There was no real pattern across ages for duration in conventional time. The difference between university students' performance and that of younger pupils was slight. Geoscience content knowledge of students in the sample varied considerably. All university students had been taught about deep time within two months prior to the interview. Yet, there were many areas in which their understanding was similar to younger pupils whose most recent formal exposure to deep time was two years prior to the interview.

What can be said is that everyone's knowledge was fragmentary. As expected, no one was able to complete all tasks successfully. Some had trouble with a number of them while others with only a few. This is not surprising if the crux of the argument is true, namely that a conception of deep time rests upon a notion of conventional time, a conception of large numbers, and relevant geoscience content knowledge. We must now consider the implications of these results and how they relate to the work of others in the field.

CHAPTER FIVE

DISCUSSION

Chapter one introduced the three research questions investigated in the interviews (p. 21), each of which corresponds to a “leg” of the “deep time stool.” In chapter four I reported the results of the interview tasks according to the three “legs” of the “stool” they represent: conventional time, large numbers, and geoscience content knowledge. This chapter is organised somewhat differently to complement chapter four and synthesise the results of the research. I attempted to show in chapter two, that in order to possess a solid concept of either conventional or deep time, a person must be able to deal with both succession and duration. A robust concept of deep time also requires an understanding large numbers and some geoscience content knowledge. Learners must apply their understanding of large numbers and geoscience content to succession or duration in a deep time context. Therefore, this chapter is organised around succession and duration. Table 5.1 lists each task in the interview protocol and whether it relates to either succession or duration. This table is a reorganisation of some of the information in Figure 3.1 (p. 170). Table 5.1 illustrates that duration tasks outnumber succession tasks in this study since duration has been explored in the literature to a far lesser extent than succession. First, I discuss the succession tasks and describe how each of the “legs” of the “stool” impact students’ understanding in that area. I then do the same thing for duration. In each instance, I connect the results of this study with pertinent research reviewed in chapter two. Due to the relative dearth of deep time literature on duration, fewer connections with previous deep time research for duration can be made than for succession. Finally, I make general comments regarding the findings of the research.

Succession tasks	Duration tasks
Fossil succession animation	Duration animations
Fossil succession static image	Duration in sedimentary layers (static image)
Card sort	Three numeric timelines
	Duration of events timeline
	Duration of events questionnaire

Table 5.1 Succession and duration tasks in interview protocol

The first research question asked whether students use the same strategies to solve problems in conventional and deep time and if they make similar errors at both time scales. I hypothesised that we would see several things if the "stool" is a useful model. First, if students lack a solid understanding of conventional time, we would anticipate similar errors on comparable tasks in conventional and deep time. Times that are outside the ability of an individual human being to experience would be just as difficult for students to deal with as those in deep time. However, it might not be possible to disentangle an understanding of conventional time from the size of the numbers involved.

The second research question asked whether students understand large numbers and proportional relationships among numbers to the extent needed to comprehend deep time. I expected at least some students to be confused about the sizes of very large numbers and the proportional relationships between them when completing tasks that required no geoscience content knowledge to accomplish.

Finally, the interviews investigated whether students cited geoscience content knowledge or everyday ideas as reasons for their answers to questions about deep time. I predicted that students who lacked the relevant geoscience content knowledge to complete the tasks would likely reason on the basis of tangential surface features. In some cases, they would get the correct answer but either the reasoning that got them there or the underlying premise upon which their reasoning was built would be faulty.

5.1. Succession

As chapter two demonstrated, succession consists of two facets. The first is the ability to sequence two or more events in temporal order, referred to as *relative time* (Trend, 2000). The second is the ability to place an event at a particular location in time. In deep time this situates an event in relation to the reference point of the present. For example, we say the end of the Cretaceous period was 65 million years ago. In a context of deep time, this is referred to as *absolute time* (see section 2.2.2). Table 5.2 lists the three succession tasks from the interviews and briefly summarises the key findings discussed in chapter four. The table does not capture the full range of results from each task, but does provide a basic summary for the reader's convenience.

Task	Key findings
Fossil succession animation	High degree of accuracy for all age groups
Fossil succession static image	Less than half the participants ordered fossils correctly. Most of the ones who ordered correctly did so based upon surface features.
Card sort	Sequencing was similar to accepted scientific sequence for events. Deep time items were no more difficult than historical ones.

Table 5.2 Summary of key findings for succession tasks

5.1.1. Succession and conventional time

The fossil sequencing task described in sections 4.1.1.1 and 4.1.1.2 assessed succession in *perceived time* and proved to be little problem for any student in this sample save one. They had no difficulty comparing the sequence of two layers and minimal difficulty sequencing the entire set of fossils in the three columns. No knowledge of fossils was required to complete the task. The fact that the picture on each layer was of a fossil was beside the point. It could have been of anything. If the students in this sample are unable to sequence a series of events in something other than perceived time it is likely not because they lack knowledge of the concept of succession per se.

There are two exceptions to that assertion. First, four students had some trouble dealing with simultaneity. Chapter four described a similar error Piaget (1969) found with much younger pupils on a task in which two lamps were turned on and off. He maintained the children were unable to distinguish their ocular movements (looking first at one lamp and then the other) from whether the lamps were lit simultaneously or successively. Thus, the error is a visual-perceptual one. The lamp I see first must have been lit first. I did not track students' ocular movements during this task, so I cannot say whether Piaget's conclusions might apply here. It is possible these students have poorer visual-spatial abilities than others in the sample. Two of the four appeared to have some dissonance between what they perceived and what they felt must be true. Both initially said the layers were successive but then decided they must have been simultaneous. The role of visual-spatial ability in this task will be discussed further when comparing this task to the work of Dodick and Orion (2003a).

The other exception was a single 8th grader who was unable to either compare the relative appearance of two fossils or sequence the entire set. It is unwise to make too much of one student's responses, since they could be an aberration. It is unclear why the task was so difficult for him. He could have visual-perceptual difficulties as his fossil order bore very little resemblance to the true sequence of the fossils. Inattention to the task is another possible explanation, although there was nothing in his demeanour to suggest that was the case. The fact that this boy was extremely confused by the task suggests that the ability to deal with succession in real time may not be universal for all young adolescents. Although admittedly a different task since it is not in real time, Friedman's (2005) findings that young adolescents could not consistently state the month that was two months prior to a specific month provide corroboration that succession is not universally mastered by this age group.

Thus, although, succession in conventional time was not a problem for most of the individuals in this sample, when students made succession errors, they were consistent with the types of errors made in studies on conventional time reviewed in section 2.2. Again, it is unwise to make much of results from one exploratory study with a small, non-representative sample. We can say that these students responded on conventional time succession tasks in ways that were consistent with what was predicted.

5.1.2 Succession and geoscience content knowledge (fossil sequencing)

A different picture emerged when students compared the appearance of two fossil layers prior to watching the animation. Here we saw clear evidence that their geoscience knowledge was a key factor in the reasoning behind their responses. Of the 15 students who said the trilobite layer formed first, all but four attributed that to something other than fossil succession. While there was no expectation that students would use the *term* fossil succession, four students demonstrated *some* understanding of the principle (see section 4.3.1). Consistent with literature reported in chapter two (e.g., Chi, Hutchinson, & Robin, 1989), many students' rationales for their answers focused on surface features. Lack of familiarity with trilobites or their physical appearance as compared to brachiopods was often used to judge age.

These results relate to Ault's (1980, 1982) work described in section 2.6.1. Children in his research focused on surface features like "crumbliness" when asked to determine the relative ages of strata in an outcrop. When a learner is asked a question but possesses minimal declarative knowledge upon which to make a judgment, decisions can only be made on the basis of surface features. If the learner doesn't know which features are important, decisions might be made based upon what appear to be the most obvious features even if they are not the most salient ones. The learner may arrive at the correct answer purely by coincidence. Given two different fossils or outcrops, the same reasoning strategy may produce an incorrect answer not a correct one. For example, chapter four showed that many participants used a familiarity strategy to determine that the trilobite was older than the brachiopod because the latter looked like a clam with which they were familiar. Yet, suppose the snail and the shark tooth fossils occupied the positions now held by the trilobite and the brachiopod. A familiarity strategy would be far less useful since species of snails and sharks are both alive today.

As described in chapters three and four, the stratigraphic sequences used for these questions and upon which the animation was based were borrowed from Puzzle 5 of the GeoTAT described by Dodick and Orion (2003a, 2003b) and used with Israeli seventh to twelfth graders. Students in their study sequenced all fossil layers from oldest to youngest with only the static image available. In contrast, students in this study sequenced all fossils only after watching the fossil sequencing animation in real time. A portion of Dodick and Orion's (2003a) results appear in Table 5.3. There were ten different fossils in the exposures, and participants were scored based upon the number of fossils ordered correctly. These scores were transferred to a percentage and mean percent correct and standard deviations were calculated. Although their sample contained students in grades 7-12, only data for grades 12, 11, and 8 appear in Table 5.3 since they are most closely related to the ages in the present sample. These authors did not report on the number of students at each grade level. Hence, that information is not included in Table 5.3.

	Grade 12		Grade 11		Grade 8	
Puzzle	Mean	SD	Mean	SD	Mean	SD
5	55.0	32.8	44.5	38.1	27.9	33.9

(Dodick & Orion, 2003a, p. 424)

Table 5.3 Mean scores and variance for three grades on the GeoTAT Puzzle 5

I calculated similar statistics for the animation sequencing task in chapter four and reported results in Table 4.1 (p. 194), which is reproduced here for the reader's convenience.

University		Grade 11		Grade 8	
Mean	SD	Mean	SD	Mean	SD
99	2.8	98	5.8	89	20.2

Table 5.4 Comparison of mean grade scores and variance for succession task, GeoTAT Puzzle 5 (N=35)

The difference between the two tasks is that succession could be observed either in real time (my task) or had to be inferred (Dodick and Orion). While Dodick and Orion report total sample size and number of classes (15), they do not break down the sample into number of students at each grade level. Hence, I have not used statistical analyses to compare their results with mine. There were no university students in their sample and no 12th graders in the present one. Results for those groups are reported for information purposes only in Tables 5.3 and 5.4. Clearer comparison can be made between the 11th and 8th graders in this sample and those in Dodick and Orion's work. In their study, students at each of those grades, on average, ordered less than half of the fossil layers correctly and variability in student responses was high. Students in my sample exhibited much greater accuracy and far less variability in their responses than in their sample which contained far more students (285). Even though standard deviations for small samples are less reliable as they frequently underestimate the true deviation, the differences between the means and standard deviations for the two samples are large enough to suggest that sample size cannot be solely responsible for the difference. The relatively larger standard deviation for the eighth graders in the present study compared to the other age groups is attributable to one very low score, but it is still lower than any of their standard deviations.

As reported in chapter two, Dodick & Orion, (2003a) attribute the poor performance of students in their sample to difficulties with temporal organisation and visual-spatial perception. They also point to the role of background knowledge in students' abilities to approach such tasks. We have already seen that some students in the present sample may have demonstrated some visual-spatial perceptual difficulties. Yet, that group represents a small fraction of the sample as a whole. Dodick and Orion are undoubtedly correct that visual-spatial perception plays a role in this task, but it may not be the main discriminator between students. In fact, a comparison of Tables 5.3 and 5.4 demonstrates that even the youngest students in the present study were able to sequence the fossils with far greater accuracy than the oldest students in their sample. The main difference was that the current task relied almost exclusively on visual perception and an understanding of succession in perceived time; the need for any geoscience content knowledge was removed. Even though articulation of the principle of fossil succession was not required, Dodick and Orion's task required at least *some* notion that two fossils of the same creature appearing in different places were cotemporaneous.

I would argue that, for the present sample, geoscience content knowledge is a significant discriminator in how adolescents or adults approach succession tasks involving geoscience events in deep time. If all a student has to go on to make a judgment is everyday experience, that experience could either help or hinder a learner's ability to sequence the fossil layers. Experience with piles of clothes in the corner of one's room (or anything else) could lead one to logically conclude that the trilobite must have preceded the ammonite in the first column since it was beneath it. Items at the bottom of an undisturbed pile were put there first. Everyday knowledge could help one correctly sequence the trilobite and the ammonite as it accords quite nicely with the geoscience principle of superposition. However if my domain knowledge is thin and I am dependent solely upon everyday experience, there is no particular reason to assume that two brachiopod layers must have formed at the same time. If I am thinking about piles of items in two children's rooms, I would not necessarily assume that a pair of black trousers found at the bottom of the pile in the first child's room was added to the pile at the same time as a similar pair in the middle of the pile in the second child's room. The difference between Dodick and Orion's results and mine are perhaps best explained by the role of geoscience content knowledge.

5.1.3 Succession and geoscience content knowledge (card sort)

The second task that explored succession in deep time was a card sort modelled after Trend (1998, 2000). There were two aspects to this item. In the first, participants sorted cards into relative sequence. Second, students were asked to place events into groups, give names for the groups, and, if possible, provide an age for each group if none had been already mentioned. Across the entire sample, there was only one difference between how participants sequenced events and the scientific consensus for their order (*1st volcanoes* and *1st life)*. Eleventh graders were the only group who placed the *origin/formation of the Sun* before the *Big Bang*. Eighth graders switched the order of two pairs of events: *1st Olympic Games* and *Julius Caesar* and *first volcanoes* and *the appearance of first life.* The latter pair was also switched by university participants.

Why might these three pairs of events create sequencing problems for students? The common denominator among them is that they are difficult to sequence in the absence of some subject matter knowledge. The statement is perhaps self-evident. Yet, it is important to note that an inability to sequence two events correctly may say very little about students' understanding of temporal order on that scale, a point made anecdotally in previous research on deep time conceptions (e.g., Ault, 1982; Dodick & Orion, 2003b; Trend, 2000, 1998). In the case of *1st Olympic Games* and *Julius Caesar*, one needs to know that ancient Greece preceded ancient Rome as a civilisation. A student may be quite capable of dealing with the temporal order of events that happened thousands of years ago. It is the events themselves that give the impression that the student does not understand. If events that were familiar to the student had been chosen, the task could have been successfully completed. As noted, 8th graders were the only group to switch the order of those two events. While they reported having studied ancient history in 6th grade, 11th graders reported studying it in both 6th and 9th grades. University students did not indicate how many times they studied ancient history in their educational careers, but all said they had done so. We have seen that number of courses taken is not a strong predictor of subject matter knowledge but it may be a factor.

To sequence *first volcanoes* and *the appearance of first life* it would be useful to know something about out gassing from volcanic eruptions and the emission of a variety of gases into the atmosphere which contain the necessary elements for organic molecules. [I have grossly simplified and skipped over steps between out gassing early in Earth's history and the origin of life since a fuller discussion is not germane to my point.] It would be helpful to know that the Big Bang had something to do with the formation of the universe to correctly sequence the *Big Bang* and the *formation of the Sun.*

However, this does not mean that students who sequenced those pairs of events accurately held scientific conceptions regarding them. In fact, it is possible to sequence the *Big Bang* and the *formation of the Sun* correctly without a clear understanding of how the Sun's formation relates to the Big Bang. While many ultimately placed the *Big Bang* before the *formation of the Sun,* they weren't universally sure if it preceded the Sun's formation or not. Some appeared to view them as occurring in very close temporal succession, if not simultaneously. Even students who placed the *Big Bang* correctly often had a poor conception of the event. They equated it with an asteroid or the formation of the Earth. Some who accurately placed *first volcanoes* before *first life* did so with seemingly little or no knowledge of evolutionary biology or Earth's early atmosphere. Rather they reasoned that volcanoes preceded life because volcanoes are an Earth process and they felt terrestrial processes take longer than biological ones. Therefore, the first volcanoes must have appeared before first life. Once again, limited domain or topic knowledge could result in a correct sequence albeit one based upon a faulty premise. Thus, the ability to place two events from the distant past in correct temporal order will reflect far more than temporal understanding alone. These results corroborate the anecdotal assertions made by others and referenced in an earlier paragraph. In some cases the failure to sequence items correctly is not specifically a temporal problem but rather evidence of insufficient knowledge of those events. Conversely, a correct sequence alone cannot be equated with temporal understanding. Additional information about *why* students ordered events the way they did is needed to determine the extent of their understanding.

Geoscience events in this study were not identical to those used by Trend for reasons that were explained in chapter three. Standard deviations for the items in this study were generally lower than those in Trend's research (2000, 2001b) although the differences are not great. Of the 20 events sequenced by U.K. primary teachers and primary teacher trainees in Trend's studies, 18 events for primary teachers and 14 for primary teacher trainees had standard deviations greater than or equal to 2.0. Of the 13 events sequenced for this study, there was only one standard deviation greater than 2.0 for the entire sample (*Big Bang),* none for 8th graders, one for 11th graders *(Big Bang),* and none for the university students.

Lack of parity between samples, specific items sequenced and research design between Trend's research and mine make it unwise to make too much of the differences between results. Differences could be due to the smaller sample size in this study. It could also have been due to the items themselves since all of his items were geoscience events while four in this study were not. The latter explanation could account for the difference in results if historical events considered part of conventional time are easier to sequence than geoscience ones. However, adjacent mean differences, standard deviations, and interview transcripts in this study indicate that students found the historical events no easier to sequence than the geoscience ones. All of these factors may have played a role in the slightly different results, but I think there is another plausible explanation.

I would argue that the main difference between his task and mine is that many of the geoscience items in the current study could be sequenced accurately using logical reasoning skills with only minimal geoscience content knowledge while he included far more geoscience items for which background knowledge in the geosciences was necessary. The logical reasoning students exhibited in this study might be based upon a faulty premise, but still produce a correct answer. For example, several people said the *formation of the Sun* preceded the *formation of the Earth* since the Sun is essential for life on Earth. That is not, however, required since the Sun *could have* captured the Earth in its orbit (even though that is not the case). It is intriguing that Trend (2000, 2001a) found the lowest standard deviations for the first three events on the list: *Big Bang, origin or formation of the Sun,* and *origin or formation of planet Earth* in his research with primary teachers and teacher trainees. I have attempted to demonstrate from my own results that students are able to sequence these events correctly even if they possess minimal or inaccurate knowledge about the events.

In contrast, the remaining items in Trend's list may be more dependent upon some geoscience content knowledge to sequence them correctly. How could one determine whether Earth's Moon was formed before or after the first volcanoes erupted on Earth without some knowledge of those events? There is nothing in the events themselves that would provide clues to their temporal order unless I know something about them. It would be very difficult to determine if the first life with hard parts appeared prior to or after the first trees appeared without knowing something about evolutionary history. In the present study, the *Big Bang* produced the highest standard deviation among 11th graders due to two students who said they had never heard of the event and placed it very late in the sequence. Since the sample was small, they had a significant effect on the standard deviation for the entire sample. That is the point. The standard deviation for that item was high because of a lack of knowledge of the event, not necessarily because of a poor understanding of deep time. *First volcanoes* produced the highest standard deviation for 8th graders and university students, and the second highest for the entire sample. While this item is difficult to place without some knowledge base, some students did so for unusual reasons. The present study corroborates Trend's assertions (1998) that the geoscience content knowledge a student brings to the table plays a crucial role in an understanding of deep time. There is likely a mutually dependent relationship here in which content knowledge and temporal understanding support the development of the other. The nature of that relationship is an unanswered question. The issue will be explored more fully in chapter six.

The addition of historical events in this study provides a benchmark against which to judge how people evaluated earlier geoscience events. Learners appeared to approach these items in exactly the same manner as the deep time events and their responses were no more accurate. In this study older students who had several exposures to ancient history expressed *less* confidence about their sequencing for the historical events than for the geoscience ones. Again, I would argue this is because earlier events in this study could be sequenced on the basis of logical reasoning strategies, while it is more difficult to sequence the historical events without some knowledge of them. This concurs with Hidalgo & Otero's (2004) contention that students in their study were able to sequence events accurately using deductive reasoning even if they lacked knowledge of the events themselves.

The types of responses seen here are highly consistent with previous research described in section 2.6.1 and support the predictions made based upon that research. There is great confusion about the temporal order of geoscience events even among students with relevant prior coursework. The events themselves seem to have little meaning for students. Absent specific knowledge, they must rely on logical reasoning strategies to determine the answer. In some cases, that works well and they come up with the correct answer. However, in another situation similar reasoning may not lead to the correct answer.

Taken together, these results indicate that the ability to comprehend succession in deep time and/or answer questions in research studies on the topic is highly sensitive to the amount of geoscience content knowledge a student brings to the table. All of the researchers with which these results were compared are well aware this is true. I would argue that it, coupled with logical reasoning ability, may be the *most* important factor in many instances, especially for adolescents or adults.

5.1.4 Succession and large numbers

The previous two sections dealt with relative placement of events. Succession also involves the absolute placement of events at a specific point in time. In-service and trainee primary teachers in Trend's research (2000, 2001a) were given temporal categories corresponding to deep time bands and asked to place the events in the appropriate category. He noted a wide range of placements with some individuals placing events in the category "more than one million million years ago," much older than the currently accepted age of the universe (Trend, 2000). In the present study, students grouped events into categories of their choosing and provided a name for their groups, if possible. While the methodology was different than his, students' conceptions of absolute time as measured in this study were quite similar to those of participants in Trend's research and in that of others discussed in chapter two (Libarkin et al., 2007; Marques & Thompson, 1997), namely that ages for piles ranged across many orders of magnitude. Age names given for the oldest group ranged from more than ten times greater than the accepted date for the Big Bang to the fourth dynasty of ancient Egypt. Nine students did not provide an age name for the oldest group. This was reminiscent of two studies reported in chapter two in which university students (Catley & Novick, 2009) or in-service teachers (Dahl et al., 2005) were resistant and, in one case, refused to provide dates for evolutionary or geoscience events.

Results from those earlier studies along with this one suggest that people simply have few age referents for events in the distant past. It is unclear if this is because very large numbers have little meaning for them or if they possess minimal knowledge of the events and couldn't place them temporally for that reason. It could be a combination of the two factors. As mentioned in chapter four, the high degree of accuracy when dating Columbus's voyage to the New World may have been largely due to a rhyme taught in American schools to help learners remember the date. In fact, a number of individuals recited the rhyme while completing the task, although not all recited it correctly. Even then, one student had the digits correct but concluded Columbus sailed to the New World in 1942.

This item suggested that age appeared to play a minimal role in students' thinking about the task. They rarely applied age names to any of the piles unless specifically asked to do so. Not only did participants not possess numerical referents for the events, they didn't even appear to view the task as a temporal one despite the fact that the directions to them stated that they should place cards into the same pile if they represented events that happened around the same time. While the task didn't provide much useful information in terms of the specific temporal categories generated, its chief usefulness was in the fact that it reinforced that students didn't appear to be thinking in temporal terms at all. One-fifth of the students were unable to provide age names for any groups.

There are several reasons why this might be true. Literature reported in chapter two indicated that people tend to perceive both numbers and time (cf. Dehaene, Izard, Spelke, & Pica, 2008 and Janssen, Chessa, & Murre, 2006) in similar ways. They have difficulty placing numbers or events accurately the farther they are from a reference point. For time, that reference point is often the present. A compression of times that are farther from the present means that I may not perceive there is a greater temporal distance between events that happened 10,000 versus 4,000 years ago than there is between ones that occurred 2,000 versus 500 years ago.

Second, how should we interpret why people assigned very old dates to the origin of the universe such as hundreds of billions of years ago? Conversely, why would another learner place the origin/formation of the Sun and Earth at 2,500 BC? Might either simply be the largest number that comes to mind at the moment? Someone may allege that the younger date is indicative of a young Earth creationist view. Yet, a person who says the Sun and Earth were formed in 2,500 BC is indicating an age that is still 1,500-5,500 years less than that espoused by people who believe the Earth is young. In some cases at least, I think a more likely explanation is that 2,500 BC merely sounds like a very long time ago to some individuals. The lack of a sufficiently large time scale on which to place geoscience events will make it difficult to develop a solid understanding of Earth's geologic history.

5.2 Duration

Duration in deep time involves the ability to judge the amount of time necessary for an event to occur or the amount of time that has passed since an event happened. Judging durations often relies on the establishment of a reference point. That could be the starting or ending time of one of the events, or it might be the present. Duration can be judged by accounting for some spatial aspect of the task—distance, size of the finished product, and the rate at which the task was completed.

When thinking about events that are not perceived in real time, a person must know something about rates of those events in order to judge durations.

Table 5.5 briefly summarises the key findings from the duration tasks in the interviews. The table does not capture the full range of results reported in chapter four, but does provide a framework for the discussion that follows.

Task	Key findings
Duration animations	On average students answered slightly more than half of the questions correctly. Lowest scores were recorded when size varied but durations were the same.
Duration in sedimentary layers (static image)	A small fraction of the sample answered correctly. Many students concluded that the thinner layer took longer to form and based their answer upon A3.
Three numeric timelines	Some students were able to deal with numbers involved in deep time while a significant number were not. Some were unable to deal with time periods up to 100 years on a linear, proportional scale.
Duration of events timeline	Events with short and long durations were estimated inaccurately. Surface features were used to determine durations.
Duration of events questionnaire	Students most accurately estimated durations for events taking seconds, minutes, or days to occur. Three of the six events that were estimated the least accurately were geoscience events.

Table 5.5 Summary of key findings for duration tasks

It will quickly become obvious to the reader that separating conventional time, large numbers, and geoscience content knowledge as factors in a conception of deep time is even more difficult for duration than it was for succession. Sections 5.2.1 and 5.2.2 contain discussions that relate to geoscience content knowledge and large numbers. To streamline the flow of the argument, I make some comments about each of those areas in sections 5.2.1 and 5.2.2. The alternative would necessitate dealing with the same data in multiple places, making the overall discussion more difficult to follow. Points made about large numbers and geoscience content knowledge in sections 5.2.1 and 5.2.2 are briefly cited in later sections.

5.2.1 Duration in conventional time via animations

Results for the three duration animations yielded interesting findings. While determining succession in real time was easy for most learners in this study, duration was not so simple. Performance when size was held constant and rate and duration varied was better than when rate and size varied but duration was held constant. While the former scenario relied on the indirect relationship between rate and duration, students did not need to judge the relative contribution of two different variables to account for duration, thus making the task easier according to Matsuda (2001). If size is held constant, I need only attend to the differences in rates to judge duration.

In contrast, the second animation in which layers were different thicknesses but had the same durations proved to be the most difficult of the three animations. To judge durations, learners needed to account for both size and rate. Based upon interview transcripts reported in section 4.1.2.2, some students in this sample were unsure how to connect those two pieces of information to answer the questions. As a result, they did not always make accurate judgments. This stands in contrast to some previous research in which adults made no errors when asked to judge durations using rate and distance (Matsuda, 2001) while supporting the findings of others that adults did not consistently judge duration accurately when considering the two factors (Casasanto & Boroditsky, 2008). The task in the present study may be more similar to Casasanto and Boroditsky's than Matsudo's. In the latter study participants watched toy trains move across a track, thus simulating an event with which individuals may be familiar. Casasanto and Boroditsky's study and the animation tasks were more abstract and any real world analogue was perhaps less apparent.

Inattention to the task was mentioned in chapter four as a possible reason for students' performance on the animation questions. Some learners chose to only watch each duration animation once even though they were given the opportunity to watch multiple times. That still does not fully account for the results reported in chapter four because participants who watched the animations multiple times also judged durations incorrectly. Several students expressed surprise that they were being asked whether a thick layer took longer to fill than a thin one for A2. Students who equated size with duration on this animation made an error similar to mistakes made by subjects in studies reported in chapter two. It is precisely the same error that young children made on Piaget's duration tasks described in section 2.4.1. However, Piaget would not have expected to see an "average adult" make this error. Yet, this is precisely the mistake Casasanto and Boroditsky (2008) found with MIT students (section 2.2.3). It is purely speculation at this point and definitely requires further research to test, but one wonders whether the human brain possesses some default notion that is applied in the absence of other information, namely that size should equate with duration. That default notion is rejected only if there is some reason to do so such as specific knowledge or another piece of perceptual information that simply cannot be ignored, such as noticeable rate differences.

Others (Dodick & Orion, 2003a) have found that students conflate size and duration in a geologic context. This is precisely the error a large number of subjects made on Puzzle 4 of the GeoTAT reported in section 2.6.2 and is where the interplay between two “legs” of the “stool” becomes difficult to ascertain. As mentioned in chapter two, this error could be seen as a lack of understanding of rates of various geological processes. In that sense there is a geoscience content knowledge component to the error. It could also be further evidence that when relevant geoscience content knowledge is missing people revert to a more basic notion of duration that equates it with size. The spatial component of an understanding of duration is powerful (Boroditsky, 2000; Casasanto & Boroditsky, 2008). When the sizes of the layers are equal as they were in A1, the spatial component has effectively been removed from the equation since it is the same for all layers. When size is different as was the case in A2, people attend to it.

The other way durations could have been judged for all three animations would have been to rely solely on the timer in the upper right corner of the screen. This necessitated attention to starting and ending times for each layer and then some one- or two-digit subtraction to determine which duration was longer. Alternatively, a person could simply count the amount of time necessary for each layer to fill either silently or aloud. While some students used a counting strategy from the beginning, others, like Hannah (section 4.1.2.4) adopted a counting strategy for later questions in the set. As we have seen, this strategy did not consistently produce correct answers. Inattention to the task may be the best explanation. While it is possible that a significant portion of the sample was unable to consistently perform the subtraction correctly, one would hope that is an unlikely explanation given the ages of the students involved. The other possibility is that given a discrepancy between perceptual information and the result of counting, the perceptual information won out (Nathan in section 4.1.2.2).

5.2.2 Application of duration to a stratigraphic sequence

The pivotal role played by perceptual information is further demonstrated by A3 and its effect on how people judged durations for the stratigraphic sequence. As described in chapter four, the most unfortunate aspect of A3 is that it may have instilled an alternative conception in the minds of some participants. While the blue layer in A3 was much thicker than the red layer, it filled at a much faster rate. Hence, its duration was shorter.   Even though size and rate both varied, the difference in rate was so significant that it made an impression on many of the participants.   When asked to judge the durations of the two sedimentary layers in the stratigraphic sequence, some learners reinterpreted the entire set of animation tasks to make them fit what they saw with the red and blue layers in A3. Several said that in each animation thinner layers filled more slowly than the thicker layers to which they compared them although that was not true. Others expressed surprise at being asked to apply information from the animations to the stratigraphic sequence.

These results can be compared with those of Dodick and Orion’s (2003a) SFT described in section 2.6.2. In chapter two, I argued that their results were best explained by the geoscience content knowledge possessed by the students in their sample. That is likely a factor, and perhaps is the most important one. However, the present results suggest that geoscience content knowledge alone may be insufficient to account for their findings. Size of a layer in the SFT was often equated with duration with a thicker layer being judged to require a longer duration than a thinner one. Thus when both outcrops had the same number of layers, but Outcrop B was taller, it was judged to have taken longer to deposit. However, students seemed to view the number of layers as a more important determiner of duration than overall size of the outcrop. When the height of the outcrops *and* the number of layers were different, the most often cited reason for judging which outcrop was older was the number of layers. As I argued in chapter two, those two ideas are not dissimilar. In both cases, a greater quantity equals a greater duration. In one case the greater quantity is the greater spatial distance occupied by the taller outcrop. In the second case, the greater quantity is the number of layers. This could well indicate a view that all depositional layers represent the same duration regardless of their thickness.

Students in this sample made errors on duration tasks in conventional time (section 5.2.1) as well as those involving deep time. They did not consistently make the same error, however. Some equated size with duration in both tasks, consistent with what was expected based upon the literature in chapter two. Conversely, the difference between the red and blue layers in Animation 3 led some to allege an inverse relationship between size and duration. In both cases, students concluded that information about the size of a layer was sufficient to determine its duration.

5.2.3 Duration and geoscience content knowledge (Timeline 3)

We have seen that geoscience content knowledge appears to be important in determining how the students in this sample make judgments about what information to attend to when determining duration. There are hints that at some individuals in this sample equated size with duration unless there was some compelling reason for them to not do so.

Students' ability to assess duration when they are provided with no perceptual information but must rely on their knowledge of events was assessed via Timeline 3. Learners had difficulty judging durations not only of geoscience events, but also events well within conventional time. Moon phases, day/night, and yearly cycles are all part of the elementary [primary] school science curriculum in the U.S. Every student in this sample had likely been taught something about each of those ideas at least once in their educational career. Nonetheless, their understandings of durations were quite inaccurate. While most knew the Earth orbits the Sun in one year, one person said the Earth rotates on its axis hourly. In some cases, their ideas regarding the durations of shorter events conformed to perceptual experience. Many participants said the Moon orbits the Earth daily which is consistent with what some undergraduates in another study thought (DeLaughter et al., 1998). In both studies, students may have reasoned that they see the Moon at night, but not during the day so its orbital period must be one day. Yet, it is highly probable that many of these same students *have* seen the Moon during the day at some point in their lifetime. The fact that the Moon is sometimes visible during the day was not pointed out to students during these interviews consistent with the aims of the study, but it would be interesting to do just that in a follow-up study to see what, if any, effect that has on students' responses.

As described in the previous section, perceived size is frequently equated with duration. This is true even if a student's knowledge of which of two structures is larger is incorrect. Several participants said the carving of the Grand Canyon required more time than the growth of a coral reef due to their belief that the canyon is larger than a reef. If size equals length, that is incorrect. While the Grand Canyon is larger than many coral reefs, the Great Barrier Reef is over six times longer than the canyon. Thus, it is possible to argue that there is at least one coral reef that is larger [longer] than the Grand Canyon. To be sure, it is quite unlikely that students actually knew the size of the Great Barrier Reef. It is also probable that at least some of them didn't know the size of the Grand Canyon either. The point is that for many students perceived size equalled duration.

Surface features other than size were also used to judge durations. The view that terrestrial processes take longer than biological ones (which also emerged on the card sort) led some students to conclude that the carving of the Grand Canyon took longer than the amount of time most coral reefs have been growing. Sometimes learners knew they did not have all the information necessary to make a judgment, but the information they said they needed to make a determination would not have helped. One university student was quite concerned that she didn't know the source of the Grand Canyon. (She did not say the source of the Colorado River) She felt that knowing the source would be important in order to judge its duration. However, knowing that the Colorado River originates in the Rocky Mountains in Colorado does not provide clues about the amount of time necessary to carve the Grand Canyon. Knowing something about how running water erodes underlying bedrock would be helpful.

Timeline 3 can be compared with Libarkin, Kurdziel, and Anderson (2007) reported in section 2.6.1. Their timelines dealt with succession while timelines in this study involved duration. The similarity is in how participants in both studies dealt with shorter versus longer time periods. In Libarkin, Kurdziel, and Anderson, many students compressed the amount of time between older events such as the appearance of first life and the appearance of dinosaurs but spaced out more recent events such as the appearance and disappearance of dinosaurs. This is very similar to what learners did in the present study. They allotted more space between events of shorter duration than was warranted but compressed events of longer duration together. Findings from both studies are consistent with studies reported in section 2.3.1 on how people map numbers (e.g., Siegler & Opfer, 2003).

5.2.4 Duration and geoscience content knowledge (duration of events questionnaire)

Just as was true for TL3, familiarity with an event was an important factor in how accurately students judged durations on the duration of events questionnaire. No one misjudged the unit of time necessary to eat dinner. This is not surprising since it is the one activity on the list that students are likely to have done practically every day of their life. As was the case for TL3, students not only misjudged durations on the questionnaire for events in deep time but also events in conventional time. In fact, the event for which duration estimates were the least accurate was the amount of time necessary for a red blood cell to make one trip throughout the body. The issue, then, cannot be that long periods of time are inherently problematic solely because they are long. On the contrary one issue is that the events involved are largely unfamiliar. (I will discuss the problem that arises because the time periods are long in the next section) For this sample, it was no easier to judge the durations of unfamiliar events in conventional time than it was to judge durations for those in deep time.

The item that produced the greatest variability in responses was the amount of time necessary for light from the Sun to reach the Earth. This was judged to be anywhere from seconds to millions of years. Again, this points out that a number of students in this sample simply have no idea about the durations of shorter events as well as those that occur in deep time. They may not know how fast light travels or what it would mean for Earth if the Sun was the right distance for its light to reach Earth in millions of years.

Results from Timeline 3 and the duration of events questionnaire lend support to the notion that geoscience content knowledge plays an important role in how students approach tasks related to deep time. It also plays an important role in how they approach tasks in conventional time. While there is nothing surprising about this assertion, it points out the need to be cautious when attempting to describe the extent of students' understanding. Perhaps if we had asked a similar question with an event with which students were familiar, we would have gotten a different response and reached a different conclusion about students' understanding.

5.2.5 Duration and large numbers

The tasks related to duration and large numbers were the most discriminable of any in the entire interview. Over half the students in this sample demonstrated sufficient difficulty with large numbers to suggest that they will find deep time challenging on that basis alone. A small subset of the sample displayed difficulty with proportional relationships between time periods up to 100 years. Perhaps more significantly, this group approached the task in a less efficient manner that almost guaranteed they would not complete it successfully. The left to right strategy they employed meant that they always ran out of room on the right side of the line. Essentially, they started at zero and compared successive pairs of numbers. Their mention of the need to show that one year is a long time could suggest that they were using themselves and their own experience as a referent in some way. When this group approached Timelines 2 and 4, their placement of numbers indicated they either understood little about the relative sizes of the numbers or attached minimal meaning to the numbers themselves.

A second group was somewhat able to place time periods up to 100 years on a proportional scale, but lost that proportionality when the numbers became larger. Their overall understanding of larger numbers was not much different from the first group. They did, however, have a better understanding of proportionality. They attempted to deal with the entire scale by placing the first and last numbers and then the remaining numbers in light of those. This stands in contrast to the first group that only dealt with two numbers at a time and not the entire scale. The second group's proportional reasoning and problem solving strategies broke down when they had to deal with unfamiliar numbers.

Students in the first two groups did not map large numbers logarithmically to the extent that would have been predicted based upon the literature reviewed in chapter two. Instead, sometimes a smaller space was left between a pair of numbers that differed by one power of ten as opposed to those that differed by two. Other times, all numbers were spaced fairly evenly across the line as if they all represented similar durations. However, logarithmic mapping was evident on TL2 in the following way. *Some* students commented that 1,000 and 100,000 and 1 million and 100 million differed by a factor of 100. Thus, they created equal spaces between the two pairs of numbers. It is entirely possible that the cognitive demands of placing seven numbers of very different durations on the same timeline were too high for some students in this sample. Many of the individuals in the first group and some in the second spaced all numbers evenly across the line for TL4.

The final group differed from the others in two ways. They were able to see the totality of the scale involved and to benchmark everything against the largest number in the set. In that sense they were quite different from the first group in that they did not use themselves as a referent but appeared able to view the entire scale from the outside as an observer. This allowed them to conceive of the proportional relationships between adjacent numbers on the required scale. It didn't mean they always got the relationships between adjacent numbers correct. This group also differed from the others in their general problem solving abilities. They saw the commonalities among the three timelines in a way that was not mentioned by students in either of the first two groups. Several of them described the second timeline as the same as the first one. Formal proportional reasoning is generally taught in the U.S. in grades 6-8 (ages 11-14) which might explain why this last group was composed primarily of 11th graders and university students. However, this group contained one 8th grader, and there were some 11th graders and university students in other groups.

5.3 Is performance in one area a good predictor of performance in another?

The answer to this question is neither simply "yes" nor simply "no." There is some slight trend in which this is true, but it seems to be more from the standpoint that some people possess better logical reasoning skills. Therefore they do better on all tasks because even if they lack background knowledge of numbers or geoscience content they are able to reason their way to many correct answers. As we have seen, however, sometimes correct answers are based on faulty reasoning strategies or premises. Students whose reasoning skills are less well-developed do not do as well on many tasks.

On the other hand, performance on the numeric timelines did not correlate well at all with performance on the duration animations. Conceptions of duration and number may not be as closely linked as succession and number appear to be. These results could be an artefact of the study's design. The data from the interviews is limited so I will not speculate too much about this point. Far more research is needed to investigate how closely related conceptions of large numbers are to both succession and duration. Possible ways to explore the connection will be discussed in chapter six.

5.4 Evidence of other factors affecting an understanding of deep time

In section 2.8 I mentioned other factors that might contribute to the stability of the deep time "stool." One of those is the credence that is attached to media representations of past events. That was evident from Michael who placed the *Pyramids of Egypt* prior to *woolly mammoths* as a result of the film *10,000 BC* in the card sort. A more recent film, *Ice Age III: Dawn of the Dinosaurs,* perpetuates the idea that woolly mammoths, sabre-tooth tigers and Tyrannosaurus-rex were all contemporaries. While many people who see the film will be aware that is erroneous, there will likely be others who will readily accept it.

The very young ages ascribed to very old events could reflect a view that says the Earth is young. Students who hold that view would obviously not suggest dates for its formation in the millions or billions. Yet, as I indicated in section 5.1.4, I do not think that can explain all of the young ages that were mentioned. Nevertheless, a young Earth view was likely demonstrated by several interview participants. Vincent, a university student, said he did not accept uniformitarianism and believed that many geologic processes do not require as much time as geoscientists think they do. While completing his card sort he said, "I disagree with the name they give it, the name Big Bang. I think it's just creation." It was not clear if Ayanna (11) was espousing a young Earth view or if she was simply aware that much of the general public in the U.S. perceives a controversy regarding the age of the Earth and the universe, although the debate many envision does not exist within the scientific community. She did, however, mention that she saw evolution (and the Big Bang) as one possible idea for the origin of the universe and that it was the only one discussed in science class.

5.5 Students' underlying conceptions

In chapter two, I discussed ways in which students' underlying knowledge within a domain can influence new learning. When a person's domain or topic knowledge is limited, the individual often relies on surface features to answer questions within the domain. There were a number of common ideas mentioned by students throughout the interviews. While these views were not mentioned by all students, they were mentioned often enough to justify discussing them. It is not clear how widespread any of these notions might be outside this sample. They could indicate further avenues for research to determine how prevalent they are and what implications they might have for geoscience educators.

The first two concern succession. One idea that appeared quite frequently in the fossil succession static image was the idea that relative age could be determined based on the familiarity of the creature. Specifically, students said if a creature looks less familiar than another creature, the unfamiliar one must be older. This was often used to deduce that the trilobite must be older than the brachiopod because the brachiopod looks similar to a clam. While students who employed this reasoning got the answer right in this instance, there are plenty of situations in which it would lead them to an incorrect response.

The second idea that emerged from the succession tasks was that if one structure (X) is dependent upon a second one (Y), then Y is older than X. Just like the other two ideas mentioned above, this one will often, but not always, be correct. Many domesticated animals are highly dependent upon their caretakers and could not survive without their care. Yet, it is clearly possible for the animal to be older than the caretaker.

The second two underlying ideas involve duration. Although it is described here under duration, the first idea surfaced in the fossil succession task as well as TL3. This view was one in which size was equated with age, or more fundamentally, duration. Students who applied this reasoning said that if one creature (or structure) is larger than another, it must be older because it takes more time for a large creature (structure) to grow (or form) than it does for a small one. Students who used this strategy in the fossil succession task relied exclusively on the size of the pictures on the image to make this judgement and not the relative sizes of the creatures themselves, presumably because they didn't know the true sizes of the creatures. This reasoning was also applied in TL3 when judging durations of events. By this logic, however, one would judge the Himalayan Mountains to be older than the Appalachians since the former are taller than the latter.

The final underlying notion that surfaced in the interviews was a view that says inorganic processes take longer than organic ones. While there are examples in which this is true, there are many examples in which this is false. An extreme example is the comparison of the Moon's orbital period around the Earth with the gestation period of a human embryo.

5.6 How do the three "legs" of the "stool" fit together?

Before answering how the "legs" of the "stool" fit together I return to the research questions outlined in section 1.7.

1. Do students apply the same strategies to solve conventional time and deep time tasks, and do they make similar errors regardless of the length of time involved?
2. Do students understand the size of numbers in the thousands or greater, as well as proportional relationships among numbers of various magnitudes?
3. When students answer questions about deep time, do they cite geoscience ideas as rationales for responses or everyday ideas that may or may not be relevant to the task at hand?

As we have seen in chapters four and five, these results suggest that students think about conventional and deep time in similar ways. The difference between them is one of degree. Students make similar errors on succession and duration tasks in both conventional and deep time. It cannot be assumed that students 13-years-old and older have an understanding of large numbers and proportional relationships among them that is adequate to enable them to comprehend deep time. In some cases, students' responses on tasks involving deep time may say more about the extent of their geoscience content knowledge than it does about any difficulty understanding very large time periods. However, this does not mean that more content knowledge will *necessarily* result in a better understanding of deep time. The picture is more complicated than that. *All* three components or "legs" of the "stool" must be present. This leads us to ask how the three factors might interact to produce a robust concept of deep time.

The "legs" of the "stool" appear to fit together in several ways. First, there is a strong spatial component to how people conceive of number and time (at least succession). The geoscience community has long recognised the role of spatial reasoning for skills like geologic mapping (e.g., King, 2008), yet I have found little discussion of its role in understanding deep time despite the fact that principles of relative dating use spatial transformations to infer temporal change (National Research Council, 2006). If geoscientists make a spatial/temporal connection, it is not surprising that learners do too. The commonality between how time and number are perceived and the implications of that commonality are key findings from this research. The reader might argue that this finding is merely an artefact of the study's design. Yet, a comparison of these results with those of others discussed in chapter two lends support to the contention that these results are not solely a result of the questions that were asked.

I believe the logarithmic mapping that occurs with both number and time happens because people do not "see" the entire scale. In order to "see" the entire scale, I must be able to mentally sit outside it as an observer and view the entire scope at once. If I am unable to do that, I am essentially viewing it from the perspective of someone who is on the scale. I may be at the origin or somewhere else on the line, but I see the scale in a manner that is analogous to the person on the road viewing two hills in the distance. The only real way to know how far apart those hills are from one another and from where we are now is to be able to sit outside the route. If I see the hills from my current vantage point I will undoubtedly misjudge the distance between them. That is one function of a map when dealing with spatial scale. If I am headed on a trip, a map allows me to sit outside the route and see the entire path on a linear scale not just what I can visually perceive from my current vantage point.

In section 2.3.1 I described a scenario in which someone might be asked to compare the difference between 5 and 7 and 55 and 57. I believe the reason why older children and adults are able to say that the numbers in each pair are the same distance apart is because they are able to step outside of the scale and look at all the numbers from the perspective of an observer. Children generally grow in this ability with numbers as they get older. Teachers of young children spend considerable time having children count objects. As a result children develop a meaningful sense of the scale in which they are working as well as numerical referents on the scale. The fact that children are able to represent a particular number in a linear fashion on one scale but cannot do so on a larger scale lends support to this assertion (Siegler & Opfer, 2003). This helps explain why large numbers are so difficult to comprehend, even for adults. We have limited experience with the numbers involved. Thus, we do not “see” the entire scale. The ability to see the scale may necessitate mentally transporting oneself to that scale. If we are unable to do that for the large end of the scale, then we can only view large quantities in relation to ourselves or the largest number that is meaningful to us. If numbers greater than 10,000 have little meaning to me, I will have severe problems comprehending the scale of deep time. This aggrees with research described in sections 2.3.2 and 2.4.2 on how experts deal with extremes of linear scale (Jones & Taylor, 2009; Tretter et al., 2006; Tretter et al., 2006).

The idea of referents, to which I already alluded, ties all three “legs” of the “stool” together. Many people do not have good referents for large quantities of any type. Powers and multiples of ten should function as those referents in the base-ten system when dealing with pure numbers. For large numbers, this requires facility with both standard and scientific notation. The problem in the U.S., at least, is that many students do not have a clear understanding of the decimal number system. The issue is further complicated for deep time because one must possess not only numerical referents but also geoscience content ones. The call for specific content reference points to help students deal with deep time has been made elsewhere (Trend, 2001b, 2001a).

This brings us to two other functions of a map for linear distances. First, it provides me with landmarks or referents along the way. If I have reached Town X, I can see from the map on a linear scale how far I still have to go. The other piece of information I get from a map is a scale of kilometres that gives me the units into which distance along my route is divided. When I reach Town X, I not only know what fraction of the trip is still to be completed but I also know precisely how much farther I must travel and can estimate how long it is likely to take. That is only true if the unit *kilometre* is meaningful to me. Many Americans travelling in other countries with a map will not find the map nearly as useful as a native because the unit into which the route has been divided (a kilometre) is not meaningful. They may remember that it is approximately 0.6 mile, but even then, they cannot "see" a kilometre in the way they can "see" a mile.

What do American travellers to foreign countries have to do with my argument? They help illustrate why scientists who work at large scales are able to make sense of the numbers and objects or events involved. First, they "see" the scale they are working in. Additionally, the units in that scale have meaning. Experts possess both numerical and content referents that serve to divide the entire scale into meaningful parts that can be combined and divided via proportional reasoning. As McPhee (1982) points out, "In geologists' own lives, the least effect of time is that they think in two languages, function on two different scales" (p. 128). The ability to function at different scales gives them an intuitive feel for the numbers in which they work (Jones & Taylor, 2009; Tretter et al., 2006; Tretter et al., 2006). People who are not experts but possess good number sense for large numbers have a portion of the tools they need to comprehend size or time at the large extreme of scale. However, the lack of content referents means they cannot place events in their appropriate context if they are not provided with relevant numerical information.

When we think about duration, the role of referents and a sense of the entire scale are more complicated. I must still be able to "see" the whole scale and have meaningful referents. However, I must also be able to use ideas about rates and perhaps size to judge duration. Possessing relevant geoscience content knowledge is perhaps more crucial than it was for succession. If I know little about average yearly rates of tectonic plate motion, I will have great difficulty imagining the amount of time necessary for the break-up of a supercontinent like Pangaea. Yet, knowledge of the event will not be sufficient by itself. I must also be able to apply multiplicative reasoning to figure out what that average yearly rate of tectonic plate motion means.

We have seen that students in this sample struggled with all three "legs" of the deep time "stool." A few performed well or poorly or all tasks, but most people were somewhere in the middle. They did well in some areas and not so well in others. The three factors investigated in this study all seem to influence how students understand deep time. None appears to be sufficient alone to account for their difficulties. All must be acknowledged for the role that they play.

CHAPTER SIX

CONCLUSIONS

This study set out to determine if the geoscience community needed a new model for how to think about students' difficulties understanding deep time. The tasks in the interviews were designed to explore the roles played by three factors in a concept of deep time. The conclusions reached point to the need to reinterpret some previous research in light of these findings.

This thesis provides an original contribution to the literature in several ways. It is the first time anyone studying the topic has explicitly tied together research literature about large numbers, conventional time, and geoscience content knowledge in the same paper (#1) and then related them to deep time (#2). Third, this study is also the first attempt to explore these three factors in the *same* investigation with the *same* individuals. Fourth, it is the first time anyone has looked at how a group of people deal with the three factors *within* and *outside* the deep time context. My exploratory data provides a warrant for the assertion that the picture is more complicated than we in the geoscience research community have conceived of it to date. I believe we need to take greater account of all three factors when trying to ascertain why students struggle so much to comprehend deep time. Finally, this study suggests that spatial skills may play a more significant role in how people understand deep time than has been previously thought. The last insight may be the most important contribution the study makes to the field.

The remainder of this chapter reviews what has been learned from the study and the ways future research might proceed in light of these findings. I begin by describing the strengths and weaknesses of the model of the deep time "stool." I then discuss the study's limitations. Suggestions for future research designed to improve upon these results are outlined. Finally, I make brief comments regarding how this study might inform the curriculum and teaching of deep time in the geoscience classroom.

6.1 Is the deep time "stool" a useful model?

The deep time "stool" was introduced in chapter two as a working model that provided a framework for this study. It has served its purpose by attempting to illustrate the role of the three factors that were explored. The study has revealed that the model possesses strengths, but also some important weaknesses.

6.1.1 Strengths of the "stool"

The construct of the deep time "stool" says that a solid understanding of deep time rests fully on the three "legs" of a person's understanding of conventional time, large numbers, and geoscience content knowledge. Instability can result if one of those "legs" is missing or is too short, although the "legs" need not be of equal length for the "stool" to be sturdy. In this sense, the model is useful.

The types of errors students make on tasks involving succession and duration in deep time mirror the errors made on those tasks in conventional time. There is nothing in these interviews that would suggest that succession and duration in deep time function in ways that are different from how they function in conventional time.

Thus, it may well be that some adolescents and adults will struggle with a concept of deep time due to difficulties with conventional time, particularly duration. This is important because it suggests we must take greater account of the research literature on how people understand conventional time. We have assumed since Ault's work (Ault, 1980, 1982) that conventional time is not a barrier for the acquisition of a concept of deep time, but my results question this view.

Secondly, nineteen of the 35 students in this sample had sufficient difficulty with large numbers to imply that the problem would hamper their ability to understand deep time. They lack good number sense for numbers in the millions and billions, and sometimes even for smaller quantities. Number sense provides the ability to see the whole scale and be able to divide it into meaningful numerical units. This impacts succession because if a student cannot conceive of a scale of 4.6 billion years, it will be difficult to place the appearance and extinction of dinosaurs on a timeline of Earth's history. Difficulty with numbers affects duration as well. If a student cannot conceptualise millions of years, that student will likely possess an unclear understanding of what it means that dinosaurs lived on Earth for around 165 million years. They will be unable to compare the amount of time dinosaurs lived to the amount of time humans have been on the planet. When the magnitude of the numbers is fuzzy, they are all just large numbers that are viewed as essentially interchangeable.

Finally, many of the findings reported in chapter four point to the role played by geoscience content knowledge in a conception of deep time. This is perhaps self-evident to any science educator or researcher operating within a constructivist framework. The important point is not merely that we know this to be true but that a student's geoscience content knowledge is a factor that we *must* take account of in our research. Although the role of geoscience content knowledge has been mentioned in previous research on deep time, this is the first study that has tried to specifically investigate its role in the concept. Being able to ascertain where understanding breaks down will help us more effectively devise instructional strategies to remediate students' difficulties. For future research, this will mean finding ways to isolate geosciece content knowledge from the other two factors to help determine its role in concept acquisition (a point that will be further developed in section 6.3).

6.1.2 Weaknesses of the "stool"

Although the "stool" has been a useful working model, it has proved to be inadequate. A satisfactory model should account for as many factors as possible that impinge upon a student's understanding of deep time. It should also elucidate relationships that may exist between factors. It should account for whether one of the factors is more foundational to a concept of deep time than the others. The "stool" does not measure up in any area. I consider the issue of relationships among factors first.

The factors appear to interact in ways that the "stool" does not capture. In the model, the three "legs" of the "stool" sit independent of one another and are not connected except by the "seat" of deep time itself. Yet, these interviews suggest there is considerable interplay among the factors. In fact, they seem to be connected in ways that go beyond deep time. Geoscience content knowledge and large numbers provide reference points around which new information can be organised. Large numbers and conventional time seem to be undergirded by similar spatial mapping strategies. In each case there is a tendency toward logarithmic mapping with an expansion of numbers or time closer to the origin and a compression the farther one goes from the origin. For time, the origin is often the present. Yet, the model of the "stool" does not illustrate the commonalities among how people map space, time, and number. Any model that accounts for difficulties with deep time probably needs to acknowledge the role played by space. This is perhaps the greatest weakness of the model—that it does not account for a factor that may be the most basic of all.

A second problem is that the "stool" does not account for whether the factors develop sequentially or simultaneously, a nontrivial matter. The literature reviewed in chapter two indicates there is a developmental component to how people understand both time and number. Whether one of those could be considered more fundamental than the other is not clear at present. It would seem logical that a conception of conventional time might be the most elementary of the three factors. This would seem to hold true for perceiving succession and duration as they occur in real time. However, it is unclear whether understanding of longer time periods within conventional time hinges upon an understanding of number, if the two ideas develop in isolation, or if they develop in tandem.

A third problem with the "stool" metaphor is that it doesn't capture the "other factors" mentioned in chapters two, four, and five. Things like metaphysical beliefs or dispositions are not represented on the "stool," yet, chapters four and five both outlined ways in which they influenced deep time conceptions for some in this sample. At least in the U.S., a student who perceives a conflict between scientific understanding of deep time and religious teachings sometimes chooses to reject the notion of deep time, even before engaging with the idea. That student may not expend the intellectual effort necessary to understand deep time, no matter what the person understands about conventional time or large numbers. The "stool's" inability to account for factors like metaphysical beliefs was apparent from the outset, but a more robust model *should* find a way to incorporate the factors.

6.1.3 Is there a better model?

I am reticent to commit to a particular model to replace the "stool" in the absence of additional data. Perhaps a rope with twisted strands that acknowledges the role played by spatial skills will prove to be a viable alternate model. This model is better able to deal with the interconnectedness of the factors than the "stool" was. It will be weaker if one of the factors is missing or of insufficient strength. It has the advantage of being able to incorporate more strands to include the "other factors" that influence a concept of deep time. Perhaps, ultimately the geoscience research community will conclude that one model is insufficient to account for an understanding of deep time. Multiple models may be necessary similar to the dual understanding of light as a wave and a particle.

6.2 Limitations of the study

As with all research, this study has limitations as well as strengths. I described general limitations of the chosen methodology in chapter three. This section deals with more specific limitations of the research.

This is a qualitative study that makes no claims regarding the representativeness of the sample chosen. It is highly probable that this sample does not represent the population of all 8th and 11th grade students or university undergraduates in the U.S., let alone around the world. The sample of 35 students is small and self-selected, which also limits the ability to generalise findings. Students at Institution A were unique amongst all members of the sample in that they were the only ones who were enrolled in a course taught by the investigator. This prior relationship could have affected their willingness to speak honestly about their ideas, since some may have been hoping to improve their course grade by their participation. While all university students in the sample were exposed to ideas about deep time in their course prior to the interview, students at Institution A may have been taught in ways that more closely mirrored the content of the interviews. Finally, their pre-existing relationship with the investigator may have enabled them to more accurately read her facial and body language than students who met the examiner for the first time during the interview. I have been careful throughout to relate the results only to the present sample and not to suggest that they are typical of the wider population. It is possible that the study did not fully capture the range of possible responses of students at these ages. Yet, there was sufficient variety in students' responses to provide a starting point for future research employing more representative sampling techniques.

Oral interviews and drawing tasks were used to collect data. Any method of data collection has trade-offs. The richness of interview data is one of its primary strengths, but there is always a danger of inconsistency or bias when interpreting oral interview transcripts. Other than the initial sorting of the numeric timelines, all data analysis was done solely by the investigator. Thus, someone might argue that the researcher's own biases were a significant factor in data interpretation. Every attempt was made to be as consistent as possible, but the danger exists. Finally, the criteria used to sort participants' responses into groups constrain the conclusions that are reached. For example, there were some "borderline" cases when sorting the numeric timelines. Had the sorting criteria been different those students may have ended up in a different category. Section 4.2.1 outlines how those decisions were made, but judgments inevitably reflect the investigator's perspective. Extensive transcripts were reproduced in chapter four to help the reader judge the validity of the criteria used to analyse data as well as the interpretations of that data and the conclusions reached.

The tasks and the interview questions themselves may not have been sensitive enough to detect students' actual conceptions. David's timelines (see section 4.2.5) are a good example. The extent of his understanding was not fully clear. In some ways, he seemed confused, but in other ways his understanding appeared to be quite sophisticated. In situations where there were questions about the extent of a student's understanding I have included specific information so the reader can evaluate the validity of my interpretations. Another concern is that the questions or probes failed to fully address Johnson and Gott's (1996) concerns about potentially leading questions. There are a few instances in which the reader could argue that leading questions were asked (see section 4.1.2.2). I reported those interchanges fully so readers could reach their own conclusions about the severity of that charge.

As described in section 3.2.2, researchers using cognitive interviews realise students do not always provide adequate reasons for why they respond in the ways they do. If students have simply said the first thing that came to mind or answered in the way they thought they were "supposed" to answer, then many of this study's conclusions are suspect. There was some triangulation of data to mitigate the criticism, but more might have done in this regard.

This study has demonstrated that, at least for this sample, understandings of conventional time, large numbers, and geosciences content *all* play a major role in an understanding of deep time, however, questions remain. This study has *not* answered the question of the relative importance of each of the factors to a concept of deep time, particularly at the individual level. In some tasks, an attempt was made to isolate the factors from one another. Overall, however, the three factors were not sufficiently disentangled to be able to comment on their relative importance for any single individual. This is especially true in tasks such as the duration of events questionnaire and the card sort. It is not clear if the three factors develop in tandem or if one or two develop first and are then followed by the third. Additionally, the relatively poor performance on the duration animations is not adequately explained. It is unclear if this was an artefact of the experimental design, a problem with the visual prompts employed, and/or their placement within the sequence of tasks. Alternatively, it could represent a genuine difficulty with duration.

The study noted but did not address any of the additional variables that may affect a concept of deep time such as metaphysical beliefs or dispositions. Their relative importance needs to be investigated.

6.3 Suggestions for future research

The questions raised by this investigation and the study's limitations suggest avenues for further research. Some future research should be descriptive and endeavour to present a fuller picture of the role being played by each of the three factors in a concept of deep time, as well as the impact of spatial skills. Intervention studies that flow out of those descriptive studies should attempt to determine ways to help students develop a better understanding of this pivotal geoscience concept. I briefly sketch out possible research in each of those areas below.

6.3.1 Additional descriptive research

First, the interview protocol itself could be modified in various ways. The length of time necessary to conduct each interview is a drawback of this protocol in that it limits access to U.S. school children. An interview that stretches across more than one class period or that appears to result in a measurable loss of instructional time can make school leaders less willing to grant access to their students. Thus, one possibility is to break the interview into several segments. This suggestion does not decrease the total loss of instructional time; however, it may be more palatable to school administrators who are concerned about stretching an interview across more than one class period. A second possibility is to revise the interview items so that students are only asked certain items if they answer others correctly. The drawback to this is that it could result in some students being misclassified. Consider David (section 4.2.5). Had he only created Timelines 1 and 2, he would have been categorised as possessing insufficient understanding of large numbers to deal with deep time. Yet, his explanation for TL4 suggests that would have been a miscategorisation.

Second, the items themselves could be modified. Results of the duration animations were the most difficult to interpret and raised the greatest number of questions of any items in the interviews. The animations should be redesigned so that there is clearly no pattern between adjacent layers (some students perceived a pattern). Future studies that use these animations should fully counterbalance the order of presentation. While A1 and A2 were counterbalanced in this study, A3 was always presented last. This may have affected outcomes. It is not clear why students were more accurate on A3 than the other two. The order of the questions should also be varied. After each animation the first question was, “Which layer took longer to fill?” The second question was “Which layer filled first?”

The fossil succession animation could be revised. One way would be to use only fossils that look similar to creatures that are currently alive. Another option would be to create a group of fictional ones. In both cases, this would reduce the likelihood that a student would reason on the basis of familiarity/unfamiliarity with the fossils to sequence them correctly. Additionally, all fossil pictures should encompass the same area to eliminate size as a factor. If these approaches were not viable, students’ responses might reveal information about additional strategies they employ to determine succession.

Number lines could be revised in several ways. First, they need to involve a greater variety of numbers with different proportional relationships between adjacent numbers. Billions were not included on any of the number lines in this study, but they should be in future research. Varying how many numbers students must deal with in each timeline could be useful to help determine whether the problems are with numbers or with cognitive load. Again, order of presentation should be counterbalanced to determine whether that has any effect upon responses.

The question of how the factors are related to one another might be addressed by adapting the interview protocol in several ways. Timelines similar to TL3 could be presented in two versions, one in which the numeric referents for the events are provided and one in which they are not. This would allow for comparison of geoscience content knowledge with knowledge of large numbers. A second place where content knowledge could be compared with large numbers is during the card sort. In the second part when students are grouping cards into piles that represent similar time periods one group could be given the numeric referents for the events while the other group would not be provided with that information. This could provide data about how people conceive of how close two events were to one another when they are provided with the dates for those events and may shed light on the extent to which difficulties are attributable to large numbers or content knowledge.

All of the adaptations to the interview protocol just described could be pursued via additional qualitative research. However, in order to have greater confidence in the generalisability of findings, mixed methods may be more useful. A quantitative questionnaire using a larger sample could employ a paper and pencil format or be adapted for computer presentation. This is a similar strategy to one being currently used by Philip Johnson in his work on children’s understanding of a substance which flows out of his earlier qualitative research on the same subject (personal communication, October, 2007). This could easily be accomplished with the existing format, although breaking administration of the questionnaire into multiple sessions would be wise. The specific adaptations to the tasks described in the previous paragraphs would strengthen a large-scale questionnaire-type study of this type. Students could still be given the opportunity to explain their answers either via a forced choice or open response format. The only caution is that some students may simply choose to say less when asked to write their answers than when asked orally. Therefore, random interviews with a subset of the sample would be important.

Section 6.4 addresses science curricular recommendations, but additional research that relates more closely to that curriculum is needed. Deep time is a part of U.S. science courses at the middle and high school levels (ages 11-18) as was shown in chapter one. There is great variability at the age at which students are taught about deep time in the U.S. Children as young as 11 could be studying the topic in sixth grade and might differ collectively from the 18-year olds learning earth science as seniors in high school. Studies employing a wider variety of age groups would be useful for several reasons. First, it might be possible to determine a lower limit for the understanding of deep time. In fact, a lower limit of grades 7-8 has been suggested by other researchers (Dodick & Orion, 2003a). If they are correct, there would be implications for when students encounter the subject in middle school. A broader study that involves more students at a greater variety of ages could assist in the development of a profile of what understanding of deep time looks like at different ages.

The other way that could be accomplished is via a longitudinal study that tracks but does not attempt to influence how students' understanding of deep time changes as they progress through school. The latter has the advantage of being able to capture the change in an individual's thinking over time. It is subject to participant attrition and the need for sustained funding. While prior research (and this study) leads one to conclude that people at a variety of ages hold similar ideas about deep time, those who do understand the concept learned it somehow. How did their understandings change and to what do they attribute the evolution of their ideas? Methodology similar to that employed by others with scale could be fruitfully adapted (Jones & Taylor, 2009). Tracking how understanding of the individual factors changes over time may reveal that a particular factor is more important at one age than another. This would provide the possibility of linking the development of a concept of deep time to a particular conceptual change theory such as Vosniadou's framework theory (Vosniadou, 1994, 2002; Vosniadou et al., 2008) that situates naive science conceptions within theories that arise from perceptual experiences. She and her colleagues (Vosniadou & Brewer, 1992; Vosniadou et al., 2004) have documented a progression in children's ideas about Earth's shape and gravity. Perhaps an analogous progression in people's understanding about deep time will be uncovered.

This ties in nicely to the current interest amongst science educators regarding learning progressions or "descriptions of the successively more sophisticated ways of thinking about a topic" (National Research Council, 2007, p. 219) as learners interact with the domain over a period of years. An entire recent issue of the *Journal of Research in Science Teaching* was devoted to this topic (Hmelo-Silver & Duncan, 2009). There are many challenges to this type of research. One is the need to adequately define the upper limit of the progression—what should students ultimately know about the topic? The Earth Science Learning Principles ("ESLI Home," n.d.) described in chapter one may prove helpful in this regard.

The final suggestion for additional descriptive research may appear to relate only tangentially to deep time. This concerns investigations into the commonalities among space, time, and number. Is there some correlation between the size of the numbers that can be dealt with at a given point, the time periods that are meaningful, and the size/spatial distance that can be conceived? Ideally, studies of this type would begin with relatively small numbers, time periods, and spatial distances or sizes. If some fairly consistent relationships among time, number, and space can be determined for smaller quantities, the methodologies could be expanded to include larger numbers, periods of time, and sizes/distances. The reason for beginning with relatively smaller quantities would be to aid in the development of the learning progression described previously. This could help establish that lower age limit at which students could reasonably be expected to think meaningfully about deep time. Multidisciplinary research that combines the expertise of geoscience and mathematics educators with that of cognitive and developmental psychologists could prove especially fruitful in the development of a deep time learning progression.

6.3.2 Intervention studies

As educational researchers, we are not merely interested in describing what is, but more importantly in providing information to practitioners that can be used to improve student learning. Intervention studies would flow nicely out of the descriptive studies just outlined. One type would be to make the interviews themselves teaching interviews. For example, the examiner could engage students in discussion about the duration animations and what they learned *before* showing them a picture of a stratigraphic sequence and asking them to compare durations of two layers. Discussion of powers of ten and their proportional relationships could be provided.

Other intervention studies would be even more closely linked to the descriptive studies described in the previous section. These could be small-scale in which a particular intervention was tested with one or two classes. For example, would direct instruction about large numbers and proportional relationships improve students' understanding of deep time? Further work would involve randomised, controlled trials which have been used with success in both the U.S. and the U.K. (e.g., National Mathematics Advisory Panel, 2008; Tymms & Coe, 2003). There are several forms these trials could take. One would be to isolate the factors one at a time. One group would be provided with instruction in a specific area: duration, succession, large numbers, or geoscience content. Pre- and post-tests would be devised and could be compared to determine whether specific instruction had an effect upon an understanding of deep time. A second type which would probably succeed the first would be to develop (or use an existing) curriculum designed to improve students' understanding of deep time and test its effectiveness via RCTs. These studies provide opportunity to test hypotheses about how the factors influence a concept of deep time. In that way, intervention studies can validate or refute the theories that have been generated in descriptive studies.

6.4 How might these results influence classroom practices?

Curricular recommendations based upon one exploratory qualitative study must be made cautiously. For the reasons outlined in section 4.5, it is impossible to say what a "typical" student at any of these ages understands about deep time based upon the current study's results. Nonetheless, if these results hold up in more extensive research, they suggest that some adolescents and adults will not be able to comprehend deep time if we do not address one or more of the underlying factors. The National Science Education Standards (section 1.4) suggest that mountain building processes be part of the middle school geoscience curriculum in the U.S. If a significant number of middle school students (ages 11-14) have difficulty dealing with deep time, this could be dealt with in one of two ways. The curriculum could be revised so that geoscience processes studied in middle school are confined to ones that encompass shorter time periods, such as volcanoes and earthquakes. Alternatively, the mountain building process could be used as a bridge to extend students' notions about time.

I can make only tentative pedagogical recommendations based upon this very limited data. First, it appears that some students (at least in this sample) need help to see that deep time and conventional time play by the same "rules." This is probably true for students at all age levels represented by the sample in this study. Perhaps explicit mention of the similarities between the two would be helpful for students. For example, when talking about rates of deposition, teachers could point out to students that big structures/events do not always represent longer time periods. Students could brainstorm a list of examples that accord with and refute this notion.

Second, science teachers may not be able to assume students come with a solid understanding of large numbers and proportional relationships from their maths classes. Based upon this sample, even some university undergraduates will have relatively poor understanding of large numbers. Geoscience teachers may need to spend time teaching about proportional relationships among large numbers if their students are to develop a good grasp of deep time.

Finally, in the U.S., early elementary (primary) classrooms often count up to and then celebrate the first 100 days of school. Elementary teachers frequently have banners that circle the classroom to count off the days and list important events throughout the year to serve as referents. A similar linear model may prove useful in geoscience classrooms to enable students to construct a time scale that is sufficiently large to deal with events in Earth's history. The widely used drawing of the geologic column in which everything prior to the Cambrian period is compressed due to space limitations may contribute to the novice learners' difficulties with the geologic time scale. The graphic presents no difficulty to those who already possess a solid conception of deep time (or those who take time to read the ages on the side) as the ages listed indicate there has been a compression. Yet, for many learners the visual image may be more powerful and give an erroneous impression about the amount of time represented by different portions of the column. Responses to the animation questions in which students ignored the timer in the corner and instead relied on perceptual information lends support to this contention. Creating a linear classroom timeline is undoubtedly more difficult to adopt at the university level where lecturers frequently share classrooms with others who teach a variety of subjects. Nonetheless, it is worth considering.

6.5 Final thoughts

The task before us is daunting. There are many questions still to be answered about how students acquire a concept of deep time and why they often do not. This line of research is worth pursuing and time well spent. Knowing how geological processes take place in the context of deep time can help us make wise policy decisions about how to manage Earth’s resources for a sustainable future.

REFERENCES

Acredolo, C., Adams, A., & Schmid, J. (1984). On the understanding of the relationships between speed, duration, and distance. *Child Development*, *55*(6), 2151-2159.

Alexander, P. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, *32*(8), 10-14.

Alexander, P., & Judy, J. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, *58*(4), 375-404.

Alexander, P., Kulikowich, J., & Schulze, S. (1994). How subject-matter knowledge affects recall and interest. *American Educational Research Journal*, *31*(2), 313-337.

American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Anderson, K., & Leinhardt, G. (2002). Maps as representations: Expert novice comparison of projection understanding. *Cognition and Instruction*, *20*(3), 283-321.

Augustine. (n.d.). *Confessions, Book XI*.

Ault, C. (1980, May). *Children's concepts about time no barrier to understanding the geologic past*. Unpublished doctoral dissertation, Cornell University.

Ault, C. (1982). Time in geological explanations as perceived by elementary-school students. *Journal of Geological Education*, *30*, 304-309.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart, and Winston.

Berndt, T., & Wood, D. (1974). The development of time concepts through conflict based on a primitive duration capacity. *Child Development*, *45*, 825-828.

Bisard, W., Aron, R., Francek, M., & Nelson, B. (1994). Assessing selected physical science and earth science misconceptions of middle school through university preservice teachers. *Journal of College Science Teaching*, *24*(1), 38-42.

Booth, J., & Siegler, R. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*(6), 189-201.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*, 1-28.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*, 579-593.

Catley, K., & Novick, L. (2009). Digging deep: Exploring college students' knowledge of macroevolutionary time. *Journal of Research in Science Teaching*, *46*(3), 311-332.

Center for Science, Mathematics, and Engineering Education. (1996). *National Science Education Standards*. Washington, D.C.: National Academy Press.

Cheek, K. A. (n.d.). A summary and analysis of twenty-seven years of geoscience conceptions research. *Journal of Geoscience Education*.

Cheek, K. A. (n.d.). Why is geologic time troublesome knowledge? In J. Meyer, R. Land, & C. Baillie (Eds.), *Threshold Concepts and Transformational Learning*. Rotterdam: Sense Publishers.

Chi, M. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In *International Handbook of Research on Conceptual Change* (pp. 61-82). New York: Routledge.

Chi, M. (2006). Two approaches to the study of experts' characteristics. In K. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 21-30). New York: Cambridge University Press.

Chi, M., Hutchinson, J., & Robin, A. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly*, *35*(1), 27-62.

Chi, M., & Roscoe, R. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering Conceptual Change: Issues in Theory and Practice* (pp. 3-27). Dordrecht: Kluwer Academic Publishers.

Chinn, C., & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1-49.

Clement, J., Brown, D., & Zietsman, A. (1989). Not all preconceptions are misconceptions: Finding 'anchoring conceptions' for grounding instruction on students' intuitions. *International Journal of Science Education*, *11*(special issue), 554-565.

Confrey, J. (1991). Learning to listen: A student's understanding of powers of ten. In E. von Glaserfeld (Ed.), *Radical Constructivism in Mathematics Education* (pp. 111-138). Netherlands: Kluwer Academic Publishers.

Cousin, G. (2007). An introduction to threshold concepts. *Planet Special Issue*, (17), 4-5.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Dahl, J., Anderson, S., & Libarkin, J. (2005). Digging into earth science: Alternative conceptions held by K-12 teachers. *Journal of Geoscience Education*, *12*, 65-68.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145-147.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*, 1217-1220.

deHevia, D., & Spelke, E. (2009). Spontaneous mapping of number and space in adults and young children. *Cognition*, *110*, 198-207.

DeLaughter, J., Stein, S., Stein, C., & Bain, K. (1998). Preconceptions abound among students in an introductory earth science course. *Eos*, *79*(36), 429-436.

Denzin, N., & Lincoln, Y. (2005). The discipline and practice of qualitative research. In N. Denzin & Y. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (3rd ed., pp. 1-32). Thousand Oaks, CA: Sage Publications.

diSessa, A. (2002). Why "conceptual ecology" is a good idea. In M. Limon & L. Mason (Eds.), *Reconsidering Conceptual Change: Issues in Theory and Practice* (pp. 29-60). Dordrecht: Kluwer Academic Publishers.

diSessa, A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy (From the "pieces" side of the fence). In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 35-60). New York: Lawrence Erlbaum Associates.

Dodick, J., & Orion, N. (2003a). Cognitive factors affecting student understanding of geologic time. *Journal of Research in Science Teaching*, *40*(4), 415-442.

Dodick, J., & Orion, N. (2003b). Measuring student understanding of geological time. *Science Education*, *87*, 708-731.

Dodick, J., & Orion, N. (2006). Building an understanding of geological time: A cognitive synthesis of the "macro" and "micro" scales of time. In C. Manduca & D. Mogk (Eds.), *Earth and Mind: How Geologists Think and Learn about the Earth* (pp. 77-94). Boulder, CO: Geological Society of America.

Dove, J. (1998). Students' alternative conceptions in Earth science: A review of research and implications for teaching and learning. *Research Papers in Education*, *13*(2), 183-201.

ESLI Home. (n.d.). . Retrieved December 19, 2008, from http://www.earthscienceliteracy.org/.

Evolution, Creationism, Intelligent Design. (n.d.). . Retrieved May 20, 2008, from http://www.gallup.com/poll/21814/Evolution-Creationism-Intelligent-Design.aspx.

Fast Facts. (n.d.). . Retrieved June 1, 2009, from http://nces.ed.gov/fastfacts/display.asp?id=31.

Food & Nutrition: National School Lunch Program. (n.d.). . Retrieved June 26, 2009, from http://www.able.state.pa.us/food_nutrition/cwp/view.asp?a=5&Q=45622.

Friedman, W. (1982). Conventional time concepts and children's structuring of time. In W. Friedman (Ed.), *The Developmental Psychology of Time* (pp. 171-208). New York: Academic Press.

Friedman, W. (1990). *About time: Inventing the fourth dimension*. Cambridge, MA: The MIT Press.

Friedman, W. (2005). Developmental and cognitive perspectives on humans' sense of the times of past and future events. *Learning and Motivation*, *36*, 145-158.

Friedman, W. (1992). Time memory and time perception. In F. Macar, V. Pouthas, & W. Friedman (Eds.), *Time, Action and Cognition: Towards Bridging the Gap* (pp. 165-172). Dordrecht: Kluwer Academic Publishers.

Goldin, G. (2000). A scientific perspective on structure, task-based interviews in mathematics education research. In A. Kelly & R. Lesh (Eds.), *Handbook of Research Design in Mathematics and Science Education* (pp. 517-545). Mahwah, NJ: Lawrence Erlbaum Associates.

Happs, J. (1984). Soil genesis and development: Views held by New Zealand students. *Journal of Geography*, *83*(4), 177-180.

Hashweh, M. (1988). Descriptive studies of students' conceptions in science. *Journal of Research in Science Teaching*, *25*(2), 121-134.

Heath, D., & Heath, C. (2009, September). The gripping statistic. *Fast Company*, (138), 59-60.

Hidalgo, A., & Otero, J. (2004). An analysis of the understanding of geological time by students at secondary and post-secondary level. *International Journal of Science Education*, *26*(7), 845-857.

Hmelo-Silver, C., & Duncan, R. G. (Eds.). (2009). Special issue: Learning progressions. *Journal of Research in Science Teaching*, *46*(6), 605-737.

Hoepfl, M. (1997). Choosing qualitative research: A primer for technology education researchers. *Journal of Technology Education*, *9*(1), 47-63.

Holyoak, K., & Mah, W. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, *14*, 328-352.

Hume, J. (1979). An understanding of geologic time. *Journal of Geological Education*, *26*, 141-143.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*, 1221-1247.

Jackson, P. W. (2006). *The chronologers' quest: The search for the age of the Earth*. New York: Cambridge University Press.

Janssen, S., Chessa, A., & Murre, J. (2006). Memory for time: How people date events. *Memory and Cognition*, *34*(1), 138-147.

Johnson, P., & Gott, R. (1996). Constructivism and evidence from children's ideas. *Science Education*, *80*(5), 561-577.

Jones, M. G., & Taylor, A. (2009). Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching*, *46*(4), 460-475.

Jones, M. G., Taylor, A., & Broadwell, B. (2009). Concepts of scale held by students with visual impairment. *Journal of Research in Science Teaching*, *46*(5), 506-519.

Jones, M. G., Tretter, T., Taylor, A., & Oppewal, T. (2008). Experienced and novice teachers' concepts of spatial scale. *International Journal of Science Education*, *30*(3), 409-429.

Kadosh, R., Tzelgov, J., & Henik, A. (2008). A synthetic walk on the mental number line: The size effect. *Cognition*, *106*, 548-557.

King, C. (2008). Geoscience education: An overview. *Studies in Science Education*, *44*(2), 187-222.

Lamon, S. (1994). Ratio and proportion: Cognitive foundations in unitizing and norming. In G. Harel & J. Confrey (Eds.), *The Development of Multiplicative Reasoning in the Learning of Mathematics* (pp. 89-120). Albany, NY: State University of New York Press.

Lancy, D. (1993). *Qualitative research in education: An introduction to the major traditions*. New York: Longman Publishing Group.

Laski, E., & Siegler, R. (2007). Is 27 a big number? Correlational and causal connections . *Child Development*, *78*(6), 1723-1743.

Levin, I. (1982). The nature and development of time concepts in children: The effects of interfering cues. In W. Friedman (Ed.), *The Developmental Psychology of Time* (pp. 47-85). New York: Academic Press.

Levin, I. (1992). The development of the concept of time in children: An integrative model. In F. Macar, V. Pouthas, & W. Friedman (Eds.), *Time, Action and Cognition Towards Bridging the Gap* (pp. 13-32). Dordrecht: Kluwer Academic Publishers.

Levin, I., Iraeli, E., & Darom, E. (1978). The development of time concepts in young children: The relations between duration and succession. *Child Development*, *49*, 755-764.

Libarkin, J., & Anderson, S. (2005). Assessment of learning in entry-level geoscience course: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education*, *53*, 394-401.

Libarkin, J., Anderson, S., Science, J. D., Beilfuss, M., & Boone, W. (2005). Qualitative analysis of college students' ideas about the Earth: Interviews and open-ended questionnaires. *Journal of Geoscience Education*, *53*(1), 17-26.

Libarkin, J., & Kurdziel, J. (2002). Research methodologies in science education: The qualitative-quantitative debate. *Journal of Geoscience Education*, *50*(1), 78-86.

Libarkin, J., Kurdziel, J., & Anderson, S. (2007). College student conceptions of geological time and the disconnect between ordering and scale. *Journal of Geoscience Education*, *55*(5), 413-422.

Liberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. *Science*, *322*, 1201-1208.

Mahoney, C. (n.d.). Part II: Chapter 3: Common qualitative methods. Retrieved August 1, 2009, from http://www.nsf.gov/pubs/1997/nsf97153/chap_3.htm.

Manduca, C., Mogk, D., & Stillings, N. (2002). *Bringing Research on Learning to the Geosciences*. Report from a workshop sponsored by the National Science Foundation and the Johnson Foundation, .

Marin, N. (2004). How can we identify replies that accurately reflect students' knowledge? A methodological proposal. *International Journal of Science Education*, *26*(4), 425-445.

Marques, L., & Thompson, D. (1997). Portuguese students' understanding at ages 10-11 and 14-15 of the origin and nature of the Earth and the development of life. *Research in Science & Technological Education*, *15*(1), 2-22.

Marshall, C., & Rossman, G. B. (2006). *Designing Qualitative Research* (4th ed.). Thousand Oaks, CA: Sage Publications.

Matsuda, F. (2001). Development of concepts of interrelationships among duration, distance, and speed. *International Journal of Behavioral Development*, *25*(5), 466-480.

McPhee, J. (1982). *Basin and range*. New York: Farrar, Straus, and Giroux.

Mertens, D. (2005). *Research and evaluation in education and psychology* (2nd ed.). London: Sage Publications.

Meyer, J., & Land, R. (2003). *Threshold concepts and Troublesome Knowledge: Linkages to Ways of Thinking and Practising within the Disciplines*. Enhancing Teaching-Learning Environments in Undergraduate Courses. Occasional, Edinburgh: School of Education, University of Edinburgh.

Monroe, J. S., & Wicander, R. (2006). *The changing Earth: Exploring geology and evolution* (4th ed.). Belmont, CA: Thomson Brooks/Cole.

Montangero, J. (1996). *Understanding changes in time*. London: Taylor and Francis.

Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Murphy, K., & Alexander, P. (2008). The role of knowledge, beliefs, and interest in the conceptual change process: A synthesis and meta-analysis of the research. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 583-616). New York: Routledge.

Myers, M., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*, 2-26.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, D.C.: U.S. Department of Education.

National Research Council. (2006). *Learning to think spatially*. Washington, D.C.: The National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, D.C.: The National Academies Press.

Nersessian, N. (2008). Mental modeling in conceptual change. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 391-416). New York: Routledge.

Orion, N., & Ault, C. (2007). Learning Earth Sciences. In S. Abell & N. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 653-687). Mahwah, NJ: Lawrence Erlbaum Associates.

Oversby, J. (1996). Knowledge of earth science and the potential for its development. *School Science Review*, *78*(283), 91-97.

Perkins, D. (2006). Constructivism and troublesome knowledge. In J. Meyer & R. Land (Eds.), *Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge* (pp. 33-47). London: Routledge.

Petcovic, H., & Libarkin, J. (2007). Research in science education: The expert-novice continuum. *Journal of Geoscience Education*, *55*(4), 333-339.

Petitto, A. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, *7*(1), 55-78.

Piaget, J. (1969). *The Child's Conception of Time*. New York: Ballantine Books.

Poduska, E., & Phillips, D. (1986). The performance of college students on Piaget-type tasks dealing with distance, time, and speed. *Journal of Research in Science Teaching*, *23*(9), 841-848.

Pons, F., & Montangero, J. (1999). Is diachronic thought a specific reasoning ability? *Swiss Journal of Psychology*, *58*(3), 191-200.

Posner, G., & Gertzog, W. (1982). The clinical interview and the measurement of conceptual change. *Science Education*, *66*(2), 195-209.

Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211-227.

Repcheck, J. (2003). *The man who found time*. Cambridge, MA: Perseus Publishing.

Ritger, S. D., & Cummins, R. H. (1991). Using student-created metaphors to comprehend geologic time. *Journal of Geological Education*, *39*, 9-11.

Rule, A. (2005). Elementary students' ideas concerning fossil fuel energy. *Journal of Geoscience Education*, *53*(3), 309-318.

Schooldigger.com -- Search and compare elementary, middle, and high schools. (2008). . Retrieved May 19, 2008, from http://www.schooldigger.com/.

Schoon, K. (1992). Students' alternative conceptions of earth and space. *Journal of Geological Education*, *40*, 209-214.

Schoon, K., & Boone, W. (1998). Self-efficacy and alternative conceptions of science of preservice elementary teachers. *Science Education*, *82*, 553-568.

Scott, P., Asoko, H., & Leach, J. (2007). Student conceptions and conceptual learning in science. In S. Abell & N. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 31-56). Mahwah, NJ: Lawrence Erlbaum Associates.

Search For Schools, Colleges and Libraries. (, n.d). . Retrieved May 19, 2008, from http://nces.ed.gov/globallocator/.

Siegler, R., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237=243.

Silverman, D. (2000). *Doing qualitative research: A practical handbook*. London: Sage Publications.

Smith, J. P., diSessa, A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115-163.

Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics intuitive rules*. Ways of Knowing in Science Series. New York: Teachers College Press.

Stavy, R., & Tirosh, D. (1999). Intuitive rules: A way to explan and predict students' reasoning. *Educational Studies in Mathematics*, *38*, 51-66.

Tourniaire, F., & Pulos, S. (1985). Proportional reasoning: A review of the literature. *Educational Studies in Mathematics*, *16*, 181-204.

Trend, R. (1998). An investigation into understanding of geological time among 10- and 11-year-old children. *International Journal of Science Education*, *20*(8), 973-988.

Trend, R. (2000). Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science. *International Journal of Science Education*, *22*(5), 539-555.

Trend, R. (2001a). An investigation into the understanding of geological time among 17-year-old students, with implications for the subject matter knowledge of future teachers. *International Research in Geographical and Environmental Education*, *10*(3), 298-321.

Trend, R. (2001b). Deep time framework: A preliminary study of U.K. primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching*, *38*(2), 191-221.

Tretter, T., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching*, *43*(3), 282-319.

Tretter, T., Jones, M. G., & Minogue, J. (2006). Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching*, *43*(10), 1061-1085.

Truscott, J., Boyle, A., Burkill, S., Libarkin, J., & Lonsdale, J. (2006). The concept of time: can it be fully realised and taught? *Planet Special Issue*, (17), 21-23.

Tymms, P., & Coe, R. (2003). Celebration of the success of distributed research with schools: The CEM Centre, Durham. *British Educational Research Journal*, *29*(5), 639-667.

Vosniadou, S. (Ed.). (2008). *International Handbook of Research on Conceptual Change*. Educational Psychology Handbook Series. New York: Routledge.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*, 45-69.

Vosniadou, S. (2002). On the nature of naive physics. In M. Limon & L. Mason (Eds.), *Reconsidering Conceptual Change* (pp. 61-76). Dordrecht: Kluwer Academic Publishers.

Vosniadou, S., & Brewer, W. (1992). Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535-585.

Vosniadou, S., Skopeliti, I., & Ikospentaki, K. (2004). Modes of knowing and ways of reasoning in elementary astronomy. *Cognitive Development*, *19*, 203-222.

Vosniadou, S., Vamakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 3-34). New York: Lawrence Erlbaum Associates.

White, R., & Gunstone, R. (2008). The conceptual change approach and the teaching of science. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 619-628). New York: Routledge.

Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology*, *13*, 231-247.

Zen, E. (2001). What is deep time and why should anyone care? *Journal of Geoscience Education*, *49*(1), 5-9.

Zuckerman, M. B. (2009). *America's Best Colleges* (2009th ed.). Washington, D.C.: U.S. News and World Report.

Appendix A

Interview Script with Correct Answers Included

Ask: How old are you? (University—What year are you?) Are you studying earth science in school now or have you studied it in the past? What topics have you learned about in earth science (geology)? Do you ever watch television programs or visit websites about earth science (geology) topics?

I will ask you some questions about earth science (geology). After you answer each question, I will ask you to tell me what you were thinking when you answered the question or what helped you to decide how to answer the way you did. I am very interested in learning about how students like you think about ideas in earth science (geology). Please answer honestly. Any information you can tell me about how you are thinking will be very helpful to me.

You may ask me any questions you would like if you think you need more information to help you answer the question. You can ask questions at any time and as many as you wish. If you are still not sure of how to respond after I've answered your questions, please say so.

I want to show you some movies. In the movies you will see some coloured layers filling up. Then I will ask you questions about the layers. You can watch each movie as many times as you like. When you think you have watched it enough to understand what is happening, I will ask you the questions.

Show animation 1 (or 2). Before asking specific questions, say—Tell me what you saw. Did the red layer or the blue layer (*point)* take longer to fill? How do you know that? **(BLUE IS CORRECT)**

Suppose that the yellow layer and the green layers had been filling at the same time in different places instead of one after the other. Which layer would have filled first? How do you know that? **(YELLOW IS CORRECT)**

Show animation 2 (or 1). Did the blue layer or the yellow layer (*point)* take longer to fill? How do you know that? **(BOTH TOOK SAME AMOUNT OF TIME IS CORRECT)**

Suppose the brown layer and the pink layer had been filling at the same time in different places instead of one after the other. Which layer would have filled first? How do you know that? **(BOTH TOOK SAME AMOUNT OF TIME IS CORRECT)**

Show animation 3. Did the green layer or the brown layer (*point)* take longer to fill? How do you know that? **(BROWN IS CORRECT)**

Suppose that the red layer and the blue layer had been filling at the same time in different places instead of one after the other. Which layer would have filled first? How do you know that? **(BLUE IS CORRECT)**

What do you know about how sedimentary rocks form? You can find sedimentary rock in places like the Grand Canyon. The movies we just saw are like models for how sedimentary rocks form. They aren't exactly the way sedimentary rocks form, but they can help us learn some things about how time is related to rock formation.

Show sedimentary layers graphic. This is a drawing of some rock layers that are found in the SW US. All of these rock layers are above ground so you can see them. Each of the designs on the drawing represents a different kind of sedimentary rock. Look at these two layers. (*Point to layers 3 & 4.)* Based upon what we just did with the movies, what can you say about which of these 2 layers took longer to form? **(IT IS IMPOSSIBLE TO TELL WHICH TOOK LONGER BASED UPON THE INFORMATION GIVEN)**

Show static image of 3 columns. Imagine that these 3 columns of rocks were discovered some distance from each other. Each of these columns has layers of rocks containing the animal fossils in the pictures. When the rock layers were formed animals that lived at that time got buried in the rock and became fossils. We will call the rock layers by the names of the fossils they contain. Geologists study the fossils found in rocks to help them figure out how old rock layers are. Look at the trilobite layer and the brachiopod layer. Which one probably formed first—the trilobite or the brachiopod? Or did they probably form at the same time? How do you know? **(TRILOBITE IS CORRECT)**

Show block animation. Now watch this movie. We will pretend we can go back in time and watch each of these layers being formed in order. You can watch the movie as many times as you want. Then I will ask you questions about it. (*after movie)* I will pick 2 layers and ask you to tell me which formed first. If you think they formed at the same time, you should say so. Which fossil layer formed first, the clam or the fish scale? How do you know? **(CLAM IS CORRECT)**

Which layer formed first, the coral or the brachiopod? How do you know? **(CORAL IS CORRECT)**

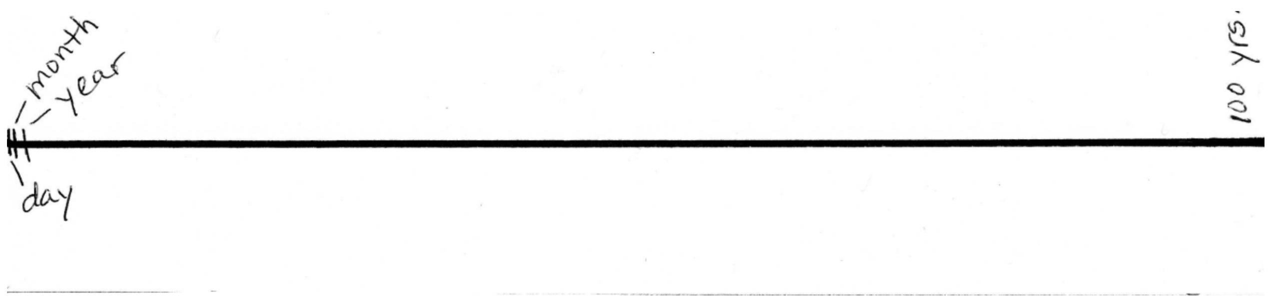
Which layer formed first, the trilobite or the brachiopod? How do you know? **(TRILOBITE IS CORRECT)**

Place each of the fossils in order from the one that was formed first to the one that was formed last. **(CORRECT ORDER: TRILOBITE, AMMONITE, CORAL, CLAM, GASTROPOD, BRACHIOPOD, FISH SCALE, SNAIL, SEA URCHIN, SHARK TOOTH)**

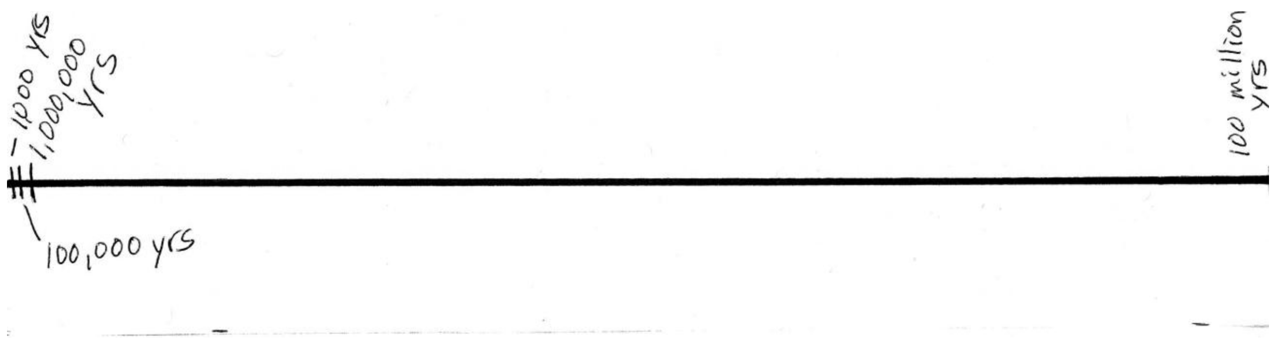
Here are some cards that list events that happened in the past. Put them in order from the one you think happened first to the one you think happened most recently. Tell me why you placed them where you did. **CORRECT ORDER: BIG BANG, ORIGIN/FORMATION OF SUN, ORIGIN/FORMATION OF EARTH, 1ST VOLCANOES DEVELOPED, 1ST LIFE APPEARED, 1ST FISH APPEARED, DINOSAURS BECAME EXTINCT, 1ST HUMANS APPEARED, WOOLY MAMMOTHS BECAME EXTINCT, GREAT PYRAMIDS BUILT, 1ST OLYMPIC GAMES, JULIUS CAESAR KILLED, CHRISTOPHER COLUMBUS SAILS)**

Now please take the cards and put them in piles based upon how long ago each of the events occurred. If two cards show events that happened around the same time, they should go in the same pile. You may make as many piles as you like. Each pile doesn't have to have the same number of cards. How did you decide where to place the cards? Please tell me a name for each of the piles. Can you put an age name on each of the piles? **(NO SINGLE "CORRECT" ANSWER FOR THIS ITEM)**

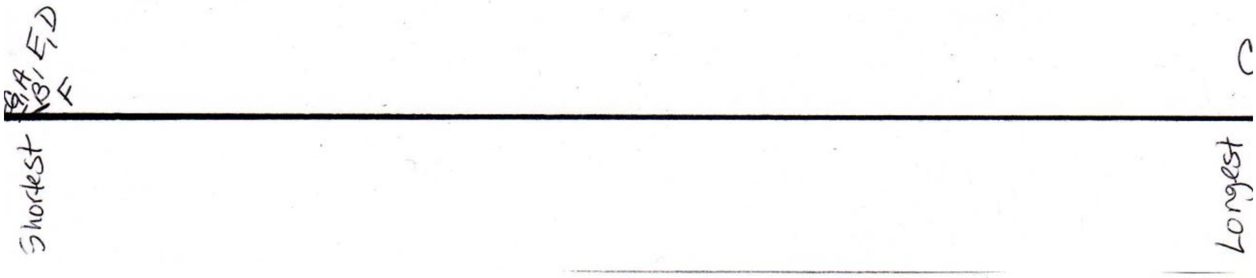
Show timeline. You've probably seen timelines before. I'd like you to make some timelines. Use this paper to make a timeline for these times (1 day, 1 month, 1 year, 100 years). Tell me about what you did. (If there's confusion show timeline I made.) Say, here's a timeline I made. I have marked 1 minute, 1 hour, and 1 day on the timeline based upon how long it takes each of them to pass in proportion to each other. **(PLAUSIBLE CORRECT TIMELINE BELOW)**



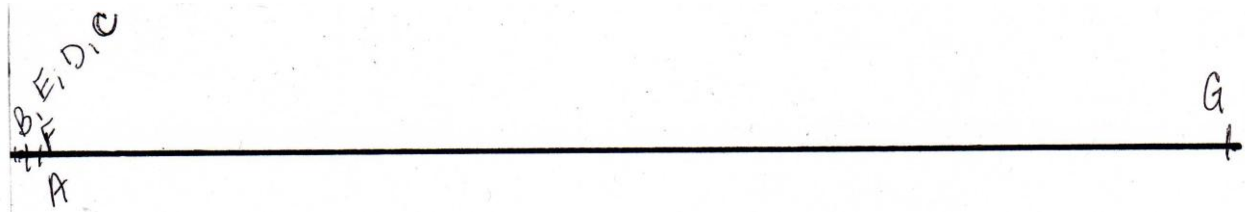
Now make another timeline for these times (1000 years, 100,000 years, 1 million years, 100 million years). Tell me how you decided where to put the numbers. **(PLAUSIBLE CORRECT TIMELINE BELOW)**



Now we will do something similar to the last question, but this time I want you to put specific events on a timeline. You know that many events occur over a period of time. For example, a soccer game takes 90 minutes (not counting half-time or injury time). The following events take place over a period of time. Put the letters in order on the line. Place the event that takes the shortest amount time on the left side of the line and the event that takes the longest amount of time on the right side of the line. Try to put the events in some kind of relative scale. For example, if one event takes about the same amount of time as another, put them close together. If one event takes a lot longer than another, put them very far apart. Tell me why you put the events where you did. **(PLAUSIBLE CORRECT TIMELINE BELOW)**



I would like you to make one more timeline. This time I wrote some times on the paper. Where would you put each of the letters now? **(PLAUSIBLE CORRECT TIMELINE BELOW)**



Let's think a bit more about how long it takes certain events to happen. Look at this table. (See next page) Place an "X" in the box that is closest to your own idea about how long each of these events takes to occur. Please place an X for each item. **(CORRECT ANSWERS INDICATED IN TABLE WITH "x")**

When finished--You may have been familiar with some of the items in the table, but you may not have been sure about some of the others. Please put a "*" next to any items that you would say you simply guessed at how long they take to occur. Tell me about how you decided where to place the X on the items you weren't sure about.

Duration of Events Questionnaire

Event	Seconds	Minutes	Hours	Days	Years	Hundreds of years	Thousands of years	Millions of years
Fly from New York to Los Angeles			X					
Colorado River to carve the Grand Canyon								X
Light to travel from the Sun to the Earth		X						
How long most coral reefs have been growing							X	
Ground to shake during an earthquake	X	X						
Spaceship to fly from Earth to the Moon				X				
Sedimentary rock to form							X	
The supercontinent Pangaea to break apart								X
A pumpkin grown from a seed to be ripe				X				
Time for Pluto to make one trip around the Sun					X			
Time it takes to eat your dinner		X						

Event	Seconds	Minutes	Hours	Days	Years	Hundreds of years	Thousands of years	Millions of years
Moon to go around the Earth once				**X**				
Build the Great Wall of China						**X**		
Appalachian Mountains to form								**X**
Hair on your head to grow ½”				**X**				
Run 100 m	**X**							
The Voyager probe to travel to Jupiter					**X**			
Drive from one side of PA to the other			**X**					
One red blood cell to travel through your whole body	**X**							
Count to one million				**X**				

Appendix B

Informed Consent for Study Participants

You are invited to take part in a study conducted by Kim Cheek as part of her PhD studies at Durham University. The study's goal is to learn more about students' understanding of geology. If you agree to take part, you will be interviewed for about 45 minutes. Your responses will be taped to allow for later analysis. When the study is finished, you will be given a written summary of the results, if you would like one.

All your responses will be anonymous. No one reading the results will be able to determine what any particular individual said in the interview. Participants like you will only be identified as undergraduate students. There are no known risks associated with participating in this research. You will be helping Kim Cheek better understand how students understand geology.

Taking part in this study is voluntary. Not taking part will not affect your grades or academic standing in any way. You are free to withdraw your consent and stop participating at any time. If you have questions at any time about the study or your role, please contact Kim Cheek, Department of Elementary and Early Childhood Education, Valley Forge Christian College at kacheek@vfcc.edu or 610-917-3936.

I have read the above consent form and agree to participate in this research.

____

____

Participant's name (please print)   date

____

Participant's signature

Appendix C

Informed Consent for Parents of Study Participants

Your child is being invited to take part in a study conducted by Kim Cheek as part of her PhD studies at Durham University. The study's goal is to learn more about students' understanding of geology. If you agree to allow your child to take part in this study and your child also agrees, s/he will be interviewed for about 45 minutes about his/her understanding about geology. The interview will be conducted at a time that is convenient for your child and his/her teachers. All responses will be taped to allow for later analysis. When the study is finished, you and your child will be given a written summary of the results, if you would like one.

All responses will be anonymous. No one reading the results will be able to determine what any particular individual said in the interview. Participants like your child will only be identified as eighth or eleventh grade students. There are no known risks associated with participation in this research. Your child will be helping Kim Cheek better understand how students understand geology.

Taking part in this study is voluntary. Not taking part will not affect your child's grades or academic standing in any way. You are free to withdraw your consent, and your child is free to stop participating at any time. If you have questions at any time about the study or your child's role, please contact Kim Cheek, Department of Elementary and Early Childhood Education, Valley Forge Christian College at kacheek@vfcc.edu or 610-917-3936.

I have read the above consent form and agree to allow my child to participate in this research.

____________________________ ___________________________ ____________

Parent's name (please print) Parent's signature date

____

Appendix D

List of University Participants

Pseudonym	Age (years)	Major
Vincent	24	Vocational ministry
Peter	19	Business
Sarah	20	Elementary (primary) education
Danielle	22	Elementary (primary) education
Cole	20	Vocational ministry
Megan	20	Vocational ministry
Claire	18	Computer science
Elizabeth	18	Computer science
Anthony	19	Geography
David	20	Geology
Lauren	20	History
Nicole	18	Undecided

Appendix E

List of 8th & 11th Grade Participants

Pseudonym	Grade	Age	Achievement Level
Evan	8	13 yrs, 11 mos.	High
Connor	8	14 yrs, 0 mos.	High
Ashley	8	13 yrs, 11 mos.	High
Emma	8	13 yrs, 9 mos.	High
Chris	8	14 yrs, 4 mos.	Middle
Kayla	8	14 yrs, 9 mos.	Middle
Alyssa	8	14 yrs, 3 mos.	Middle
Matt	8	14 yrs, 6 mos.	Middle
Ben	8	14 yrs, 11 mos.	Low
Jamal	8	14 yrs, 5 mos.	Low
Jenna	8	14 yrs, 9 mos.	Low
Sofia	8	16 yrs, 1 mo.	Low
Sean	11	17 yrs, 8 mos.	High
James	11	17 yrs, 1 mo.	High
Ayanna	11	16 yrs, 1 mos.	High
Justin	11	17 yrs, 2 mos.	Middle
Ryan	11	17 yrs, 5 mos.	Middle
Nathan	11	17 yrs, 11 mos.	Middle
Hannah	11	16 yrs, 9 mos.	Middle
Leah	11	16 yrs, 8 mos.	Middle
Michael	11	17 yrs, 6 mos.	Low
Malik	11	16 yrs, 3 mos.	Low
Vanessa	11	16 yrs, 9 mos.	Low